

Grau en Estadística

Títol: Prognosi del Gran Premi de Formula 1 en el Circuit de Catalunya de l'any 2017

Autor: Sergio Álvarez Berezo

Director: Josep Anton Sánchez Espigares

Departament: Estadística i Investigació Operativa, UPC

Convocatòria: Juny 2018



UNIVERSITAT DE
BARCELONA



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat de Matemàtiques i Estadística

RESUM

L'objectiu principal d'aquest treball consisteix a intentar predir el resultat final de la cursa de la Formula 1 del Gran Premi de Catalunya, realitzada al circuit de Montmeló a l'any 2017. Per tal de poder complir aquest objectiu, s'ha fet un anàlisi de les sessions de pràctiques prèvies a la cursa. La base de dades utilitzada s'ha construït a partir de dos documents diferents: un document que mostra els temps de les voltes realitzades i un altre que indica els tipus de pneumàtics emprats a cada sessió de pràctiques.

La part més important d'aquest l'estudi recau sobre una exhaustiva depuració de les dades per tal de poder detectar les simulacions de classificació i de cursa, amb la posterior construcció d'un model predictiu que pugui ajudar a saber el resultat final a partir de les variables explicatives de la base de dades. Finalment, també s'ha necessitat utilitzar una base de dades amb un llistat de les voltes de les pràctiques del Gran Premi de Catalunya de l'any 2016, i així poder comprovar que els criteris escollits per a la identificació de la simulació de cursa són vàlids per a dades semblants.

D'aquesta manera, el punt més important és comprovar que les conclusions finals tenen un sentit coherent i presenten certa concordança amb el resultat real.

Classificació AMS: 62J05 Linear regression, 62F03 Hypothesis Testing, 62P35 Application to physics

SUMMARY

The main objective is to try to predict the final result of the Formula 1 race of the Grand Prix of Catalonia, at the Montmeló circuit in 2017. In order to be able to carry out this objective, an analysis has been made of the pre-race practice sessions. The database has been constructed from two different documents: a document that shows the times of the laps and another that indicates the types of tires used in each practice session.

The most important part of this study is based on an exhaustive depuration of the data and to be able to identify the classification and race simulations, with a later construction of a predictive model that can help to know the final result from the explanatory variables of the database. Finally, it was also necessary to use another database with a list of the laps of the practices of the Grand Prix of Catalonia in 2016, and thus to be able to verify that the criteria chosen for the identification of the race simulation are valid for similar data.

What interests us, then, is to see if the final conclusions make sense and agree with the actual result.

Classification: 62J05 Linear regression, 62F03 Hypothesis Testing, 62P35 Application to physics

INDEX

GLOSARI	1
1. INTRODUCCIÓ.....	2
1.1 ENTRENAMENTS LLIURES	6
1.2 CLASSIFICACIÓ	6
1.3 CURSA.....	7
1.4 PUNTUACIÓ	8
1.5 OBJECTIUS	9
2. METODOLOGIA.....	10
2.1 INTRODUCCIÓ.....	10
2.2 FORMULACIÓ.....	11
2.3 VALIDACIÓ DEL MODEL.....	16
2.4 PLOTLY.....	16
2.5 FORMATTABLE.....	19
3. CONSTRUCCIÓ DE LA BASE DE DADES	20
4. LES DADES	24
4.1 SIMULACIÓ DE CLASSIFICACIÓ.....	26
4.2 SIMULACIÓ DE CURSA.....	30
4.3 EVOLUCIÓ DE LA CURSA.....	40
4.4 COMPARACIÓ DELS RESULTATS AMB LA REALITAT.....	42
5. CONCLUSIONS	49
6. ANNEX.....	52
Annex A: GP 2017	52
Annex B: GP 2016	74
Annex C: Codi	94

TAULES

Taula 1: Equips del Campionat	5
Taula 2: Sessions del cap de setmana del Gran Premi	6
Taula 3: Punts pel campionat segons la posició final	8
Taula 4: Pilots titulars de l'equip RenaultRS.....	24
Taula 5: Estimació de la evolució dels temps en la cursa per a les voltes 2,10 i 64	40
Taula 6: Evolució real dels temps en la cursa per a les voltes 2,10 i 64	41
Taula 7: Ordre de la graella de sortida real respecte estimat	42
Taula 8: Estimació de l'ordre final de la cursa (sense incidents)	46
Taula 9: Estimació del resultat final de la cursa indicant incidents	46
Taula 10: Estimació del resultat final de la cursa corregit per alguns incidents.....	47
Taula 11: Resultat final real de la cursa.....	47

FIGURES

Figura 1: Situació inicial de la cursa.....	17
Figura 2: Situació final de la cursa	18
Figura 3: "Autoescale" de la situació final de la cursa.....	18
Figura 4: Dades en format Excel.....	21
Figura 5: Dades en format Excel amb el compost utilitzat en cada volta.....	22
Figura 6: Base de dades definitiva.....	22
Figura 7: Freqüència dels temps	25
Figura 8: Freqüència dels temps per pilot.....	25
Figura 9: Voltes ràpides per compost i pilot.....	26
Figura 10: Temps per sessió	27
Figura 11: Millor temps de les FP per cada pilot	27
Figura 12: Temps de S.Vettel a la FP2 per compost	30
Figura 13: Resultat del Model (m3).....	36
Figura 14: Rendiment i degradació segons compost i any	37

Figura 15: Efecte aleatori dels pilots	38
Figura 16: Efecte aleatori dels equips	39
Figura 17: Diferències entre la volta de classificació i la volta més ràpida de les pràctiques .	43
Figura 18: Correlació entre FPBestLap i FinalLap	44
Figura 19: Correlació entre pred i FinalLap	45

IMATGES

Imatge 1: Mapa del circuit de Catalunya, Montmeló.....	3
Imatge 2: Pneumàtics produïts per Pirelli	4
Imatge 3: Compostos escollits per el GP de Catalunya	4
Imatge 4: Voltes de S.Vettel a FP3	20
Imatge 5: Resum de la FP3	21
Imatge 6: Comparació aleró posterior amb poc i molt drag	29

GLOSARI

Aturada en boxes o pit stop: és el terme utilitzat en les carreres de motor per al·ludir a la parada que fa un vehicle per canviar els pneumàtics, fer reparacions o ajustos mecànics.

Drag, arrossegar en anglès: és una paraula que s'utilitza habitualment per referir-se a la resistència aerodinàmica que ofereix el monoplaça. Un cotxe amb molt de *drag* ofereix molta resistència aerodinàmica.

DRS: sigles de '*Drag Reduction System*' és un dispositiu que té el propòsit de reduir la càrrega aerodinàmica de l'aleró posterior del monoplaça en moments determinats i així augmentar la seva velocitat punta per facilitar els avançaments.

Els flaps: aquests estan situats al aleró de davant i del darrere. Serveixen per augmentar la càrrega aerodinàmica quan estan en una posició més vertical o per augmentar la velocitat punta quan estan en una posició més horitzontal.

FP1, FP2 i FP3: sigles de *free practice* , fa referència als entrenaments lliures.

Halo: és una gàbia metàl·lica insertada al monoplaça que protegeix el cap del pilot en els accidents.

La pole position: és el nom que rep la primera posició de la graella de sortida.

El safety car, cotxe de seguretat en anglès: aquest s'utilitza per neutralitzar la cursa en el cas en el que córrer no fos segur. Tots els pilots s'han de col·locar darrere d'aquest amb cap possibilitat d'avançar fins que el director de cursa consideri que el circuit compleix en condicions de seguretat necessàries per a la repesa de la cursa.

Un set: és un joc de pneumàtics i aquest està format per quatre rodes.

El set up: és la posta a punt que fa el pilot del seu monoplaça.

Undercut: consisteix en parar a boxes per canviar els pneumàtics abans que el cotxe que tens endavant, per després fer una volta amb les gomes noves i, que quan ell entri després de fer una volta més amb les gomes gastades, en sortir de la seva parada, acabi darrere teu.

El virtual safety car , cotxe de seguretat virtual en anglès: és un sistema amb una finalitat molt semblant a les del safety car. A diferència del safety car aquest no agrupa a tots els cotxes sinó que es mantenen les diferències entre pilots, ja que tots han de seguir una velocitat predeterminada.

Xassís: consisteix en una estructura interna que sosté, aporta rigidesa i dóna forma a un vehicle.

1. INTRODUCCIÓ

La Fórmula 1, abreviada com a F1, tracta d'una competició d'automòbils que es va iniciar l'any 1950. És considerada per una majoria com la competició més prestigiosa del món del motor de les quatre rodes. Part de la seva espectacularitat prové de la velocitat punta dels seus cotxes, que pot arribar a ser major de 360 km/h, i també de l'elevada velocitat de pas pels revolts. Tanmateix, el rendiment del cotxe a la cursa no només depèn del pilot o del motor, sinó que també intervenen altres factors com l'electrònica, l'aerodinàmica, la suspensió i l'estratègia de pneumàtics. Això pot arribar a fer que, a priori, un pilot que no sigui a l'equip adequat en el moment oportú, no tingui gaires oportunitats per a guanyar el campionat.

La reglamentació de la Fórmula 1 sofreix canvis cada any i, en bona part, aquests canvis són deguts a temes relacionats amb la seguretat, com la implementació del *halo*¹ per la temporada 2018, però també hi ha canvis amb la intenció de millorar l'espectacularitat de l'esport. Per exemple, a l'any 2017, es va canviar radicalment la influència de l'aerodinàmica en el cotxe, a més d'una variació substancial de l'amplada de les rodes. Aquests dos canvis sumats implicaven que els cotxes fossin entre dos i quatre segons més ràpids per volta depenent del circuit i les circumstàncies. Una xifra molt elevada en un esport en què una dècima de segon és la diferència entre guanyar o perdre.

Com era d'esperar; amb aquests canvis els cotxes eren més ràpids als revolts, però el fet de tenir uns alerons més amples tant al davant com al darrere i unes rodes amb major superfície, implicava inequívocament una pèrdua de temps a les rectes per la resistència aerodinàmica. Aquesta modificació de les normes també va suposar una pèrdua de temps a *l'aturada en boxes*², ja que en ser més ampla la roda i, per tant, més pesada, feia que el canvi de rodes fos més lent que els de l'any anterior.

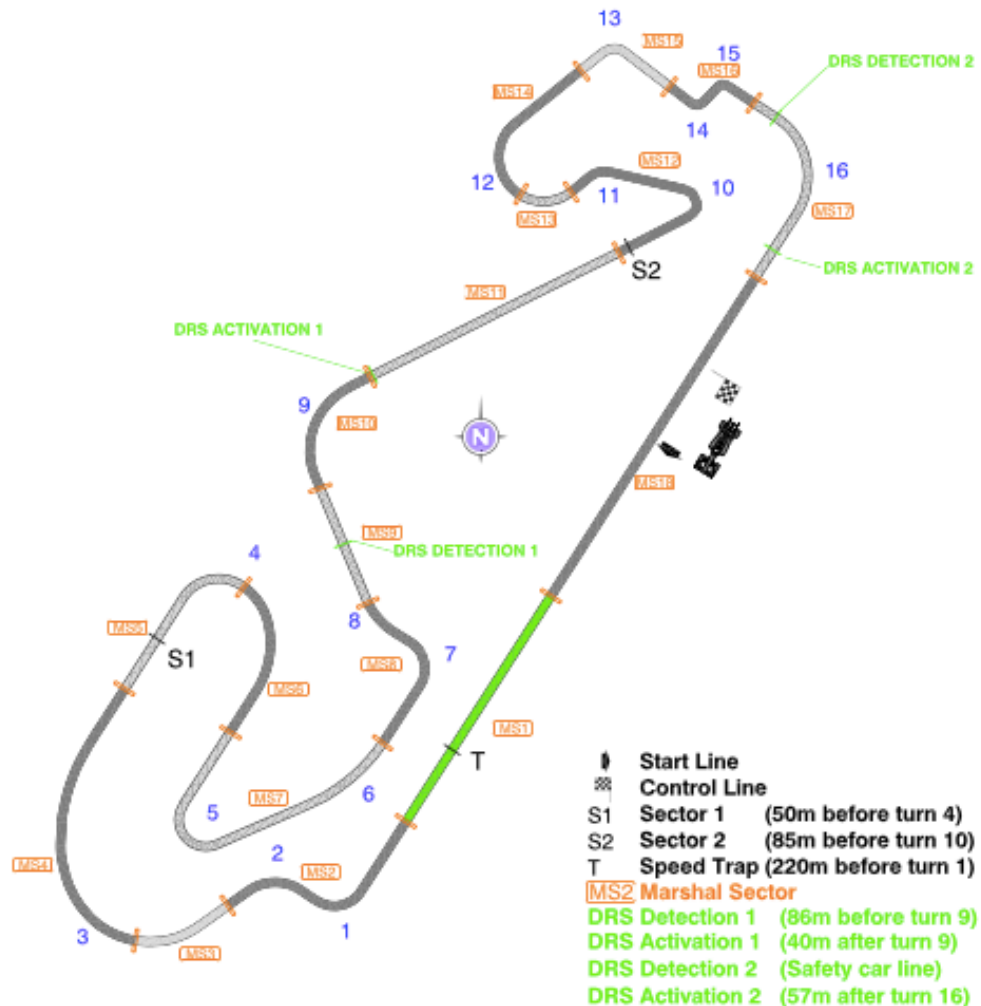
Tots aquests canvis provoquen moltes despeses en investigació per als equips i, sobretot, al tractar-se d'una competició de tan alt nivell, la despesa és molt elevada. És per això que molts equips s'han anat reduint paulatinament o fins i tot desapareixent, encara que també han sorgit d'altres nous. Molts equips han participat en la F1, però l'únic que sempre ha estat des dels inicis és l'equip Ferrari, el més prestigiós de tots, seguit de l'equip McLaren i Williams. Aquests tres equips són els considerats "històrics" a la Fórmula 1.

¹ El *halo* és una gàbia metàl·lica insertada al monoplaça que protegeix el cap del pilot en els accidents.

² *L'aturada en boxes* és el terme utilitzat en les carreres de motor per al·ludir a la parada que fa un vehicle per canviar els pneumàtics, fer reparacions o ajustos mecànics.

Pel que fa a la temporada 2017, hi ha: 20 pilots, 10 equips, 4 motoristes i un subministrador de pneumàtics, Pirelli, qui en cada gran premi porta 3 tipus de compostos per a tots els equips i només es diferencien pel rendiment i la durabilitat.

IMATGE 1: MAPA DEL CIRCUIT DE CATALUNYA, MONTMELÓ



Font: PDF de la web oficial de la Formula 1

El circuit utilitzat per realitzar la prognosi, és el circuit de Catalunya (Imatge 1: Mapa del circuit de Catalunya, Montmeló) situat concretament al municipi de Montmeló. El traçat té 16 revolts i dues rectes en les quals es pot utilitzar el *DRS*³. La cursa de Montmeló és la primera que se celebra a territori europeu i la cinquena del calendari, just després de Sochi, Rússia.

Per a aquesta cursa, tots els equips porten millores i evolucions aerodinàmiques als monoplaques, ja que en curses prèvies han tingut l'oportunitat de comprovar si hi ha

³ El *DRS* (en anglès 'Drag Reduction System') és un dispositiu que té el propòsit de reduir la càrrega aerodinàmica de l'aleró posterior del monoplaça en moments determinats i així augmentar la seva velocitat punta per facilitar els avançaments.

correlació entre les dades obtingudes amb cotxe a pista i el túnel de vent en el qual s'han provat prèviament aquestes peces. Amb tot això i, a més, per la proximitat a les fàbriques dels equips i pels entrenaments de pretemporada que únicament es realitzen en aquest circuit, la majoria d'equips decideixen portar moltes evolucions i veure quant ha millorat realment respecte la pretemporada.

Aquest circuit és un dels millors per veure com funciona el teu cotxe en termes aerodinàmics. És un traçat que aconsegueix ressaltar els defectes i les virtuts del monoplaça: disposa de grans rectes, revolts ràpids (la majoria situats en el primer i segon sector del circuit) i lents (a partir de la zona de l'estadi al inici del sector tres, que comença al revolt 10), per això, és també que es fan tants entrenaments des de fa tants anys.

Els pneumàtics produïts per Pirelli (Imatge 2: Pneumàtics produïts per Pirelli) per córrer tota la temporada són, d'esquerra a dreta: Intermedis de pluja, pluja extrema, Super Soft (Súper Tou), Soft (Tou), Medium (Mitjà) i Hard (Dur).

IMATGE 2: PNEUMÀTICS PRODUÏTS PER PIRELLI



Font: <https://www.pirelli.com/tyres/es-es/competition/f1/gama-neumaticos-f1>

Entre tots els pneumàtics que hi ha, els de pluja sempre estan disponibles per si fossin necessaris. Però de la resta, Pirelli ha d'escollir-ne tres (Imatge 3: Compostos escollits per el GP de Catalunya). Els pneumàtics escollits per córrer el gran premi van ser: Tou (Groc), Mitjà (Blanc) i Dur (Taronja).

IMATGE 3: COMPOSTOS ESCOLLITS PER EL GP DE CATALUNYA













Font: <http://www.caranddriver.es/formula-1/pirelli-llevara-al-gp-de-espana-2016-el-blando-el-medio-y-el-duro>

Els equips han d'escollir el número de pneumàtics que volen de cada compost dels que Pirelli ha escollit prèviament. Com a mínim han d'agafar un *set*⁴ de cada un.

A continuació (Taula 1: Equips del Campionat), es mostren els equips que van participar al campionat de Fórmula, a la temporada 2017:

TAULA 1: EQUIPS DEL CAMPIONAT

MONOPLAÇA	EQUIP	MOTOR	PILOTS
	McLaren	Honda	F.Alonso S.Vandoorne
	ScuderiaFerrari	Ferrari	S.Vettel K.Raikkonen
	RedBull	Renault	D.Ricciardo M.Verstappen
	Haas	Ferrari	R.Grosjean K.Magnussen
	ForceIndia	Mercedes	E.Ocon S.Pérez
	ToroRosso	Renault	C.Sainz D.Kvyat
	MercedesAMG	Mercedes	V.Bottas L.Hamilton
	Williams	Mercedes	L.Stroll F.Massa
	Sauber	Ferrari	P.Wehrlein M.Ericsson
	Renault RS	Renault	J.Palmer N.Hulkenberg

Font: <https://fr.motorsport.com/topic/guide-f1-2017/>

Tots aquests cotxes estan obligats a incloure un motor híbrid, és a dir, un motor tèrmic convencional ajudat per un altre elèctric que carrega les bateries gràcies a l'energia que es produeix a les frenades.

⁴ Un *set* és un joc de pneumàtics i aquest està format per quatre rodes.

L'estructura de les sessions que realitzen durant el gran premi és la següent:

TAULA 2: SESSIONS DEL CAP DE SETMANA DEL GRAN PREMI

SESSIÓ	HORARI	DURADA	DIA
Entrenaments lliures 1 (FP1)	10:00 – 11:30	90 minuts	divendres
Entrenaments lliures 2 (FP2)	14:00 – 15:30	90 minuts	divendres
Entrenaments lliures 3 (FP3)	11:00 – 12:00	60 minuts	dissabte
Classificació	14:00 – 15:00	60 minuts	dissabte
Cursa	14:00 – Aprox. (15:30)	60 - 90 minuts (Max. 2 hores)	diumenge

Aquesta taula (Taula 2: Sessions del cap de setmana del Gran Premi) es manté aproximadament constant en les curses europees, però moltes varien, no només pel canvi horari, sinó que a més hi ha tres curses nocturnes durant tota la temporada, on il·luminen els circuits amb milers de focus.

1.1 ENTRENAMENTS LLIURES

Durant els entrenaments lliures, els equips tenen llibertat de poder sortir a pista en el moment que creguin més oportú durant el temps que dura la sessió. Poden provar qualsevol configuració del cotxe i el compost de pneumàtic que vulguin dels tres escollits pel gran premi. Disposen de quatre hores de proves repartides en tres sessions, abans que comenci la classificació. En el moment que s'acaben els entrenaments, no es poden fer cap tipus de modificacions tècniques als monoplaques i entren en un parc tancat per evitar modificacions de *set up*⁵. Només es permet modificar la inclinació dels *flaps*⁶ de l'aleró de davant tant a la classificació com a la cursa.

1.2 CLASSIFICACIÓ

Per decidir l'ordre de la graella de sortida s'utilitza un sistema de classificació per eliminació. Consisteix en la celebració de tres sessions denominades Q1, Q2 i Q3, en les que es van desqualificant pilots de manera progressiva i en les que no hi ha límit de voltes:

⁵ El *set up* és la posta a punt que fa el pilot del seu monoplaça tant de l'equilibri, com el balanç de pesos, com les suspensions.

⁶ Els *flaps* serveixen per augmentar la càrrega aerodinàmica quan estan en una posició més vertical o per augmentar la velocitat punta quan estan en una posició més horitzontal.

- La primera sessió, o Q1, dura 18 minuts i s'eliminen els 5 pilots més lents, els quals sortiran en la posició en la que han acabat en aquesta sessió, i podent haver triat el compost amb el qual començar el gran premi.
- La segona sessió, o Q2, dura 15 minuts i es tornen a eliminar els 5 pilots més lents, seguint el mateix procediment que en la Q1, triant també el compost que utilitzaran. Els eliminats sortiran a les posicions obtingudes. Els 10 pilots més ràpids que passen a la Q3 hauran de començar la cursa del gran premi amb el mateix pneumàtic amb el qual han fet el temps més ràpid de la Q2.
- La tercera i última sessió, o Q3, dura 12 minuts i és on competeixen els 10 pilots restants, comptant els seus temps per a la graella final. Aquests pilots seran els únics que podran lluitar per la *pole position*⁷.

És important destacar que, entre les diverses sessions, hi ha una pausa de 5 minuts perquè els pilots puguin tornar al box per canviar els pneumàtics i omplir el dipòsit de combustible, en cas que fos necessari.

1.3 CURSA

Pel que fa a la cursa, tots els pilots poden sortir del box trenta minuts abans de la sortida i donar les voltes que vulguin al circuit, però han de tenir en compte que han d'arribar a la seva posició de la graella quinze minuts abans de començar la cursa. Una vegada posicionats tots els participants de la graella de sortida, com hem dit abans, els deu últims pilots poden escollir el compost de pneumàtic que vulguin dels tres que tenen disponibles. Respecte als altres deu primers, han de sortir amb el pneumàtic amb el qual van fer la volta més ràpida a la Q2. Normalment surten amb el compost més ràpid, ja que aporta una millor tracció en el moment de la sortida. Una vegada la cursa ja ha començat, els pilots han de passar com a mínim una vegada per boxes per canviar els pneumàtics i com a mínim han d'utilitzar dos compostos diferents dels tres que disposen els equips. Suposant que un pilot fa la primera parada a boxes i al canviar de pneumàtics utilitza de nou el mateix compost que portava, aquest pilot haurà de parar, com a mínim, una altra vegada més per complir la norma. En aquestes parades a boxes només es canvien les rodes, ja que els cotxes, per normativa, comencen la cursa amb els dipòsits de combustible plens i no poden parar a omplir-los de nou.

⁷ La *pole position* és el nom que rep la primera posició de la graella de sortida.

En el cas en el que en algun període de la cursa es produeixi una situació perillosa pels pilots, ja sigui per un accident o perquè la pista estigués bruta, entre altres, el director de cursa pot utilitzar diverses eines per alertar i/o solucionar el problema. Aquest, pot utilitzar unes banderes grogues, ondejades per els comissaris de pista que es troben al voltant del circuit per senyalitzar incidents menors o sinó la utilització el *safety car*⁸ o *virtual safety car*⁹ per a incidents més greus. Durant aquests períodes extraordinaris de bandera groga, *safety car* o *virtual safety car* els pilots no poden avançar-se entre ells i tampoc poden marcar temps competitiu.

Finalment el pilot guanyador serà el que completi més ràpidament les voltes estipulades en el gran premi sempre que la duració de la cursa no superi les dues hores. En aquest cas, en el moment en el que es compleixi el temps màxim establert, només haurà de completar una única volta més per donar per finalitzada la cursa.

1.4 PUNTUACIÓ

Al acabar el gran premi i tenint en compte la posició final en la cursa, es reparteixen els punts per el campionat de pilots i d'equips (Taula 3: Punts pel campionat segons la posició final):

TAULA 3: PUNTS PEL CAMPIONAT SEGONS LA POSICIÓ FINAL

1º	25 punts	6º	8 punts
2º	18 punts	7º	6 punts
3º	15 punts	8º	4 punts
4º	12 punts	9º	2 punts
5º	10 punts	10º	1 punt

La puntuació que obtingui un equip és la suma dels punts aconseguits pels dos pilots. Al finalitzar les vint-i-una curses que componen el campionat, es reparteixen dos premis: el millor pilot i el millor equip. Per tant, els guanyadors són els que aconseguixin sumar més

⁸ El *safety car*, cotxe de seguretat en anglès, aquest s'utilitza per neutralitzar la cursa en el cas en el que córrer no fos segur. Tots els pilots s'han de col·locar darrere d'aquest amb cap possibilitat d'avançar fins que el director de cursa consideri que el circuit compleix en condicions de seguretat necessàries per a la repesa de la cursa..

⁹ El *virtual safety car*, cotxe de seguretat virtual en anglès, és un sistema amb una finalitat molt semblant a les del safety car. A diferència del safety car aquest no agrupa a tots els cotxes sinó que es mantenen les diferències entre pilots, ja que tots han de seguir una velocitat predeterminada.

punts durant tota la temporada i acabaran sent proclamats els campions de Formula 1 d'aquell any.

1.5 OBJECTIUS

El principal objectiu de l'estudi és fer una prognosi del resultat final a partir de la informació extreta dels temps de les tres sessions de pràctiques. Per tant, amb aquestes dades és pretén identificar les simulacions de cursa realitzades per cada pilot, i així, crear un model estadístic que sigui capaç de predir el ritme de cursa de cada monoplaça i anticipar l'ordre final.

Entre els objectius secundaris, es troba la identificació de la simulació de classificació, també la categorització dels pilots i els equips de la graella i finalment calcular la millor estratègia de parades en boxes a partir del nivell de degradació presentada pels pneumàtics, ja que és obligatori canviar el compost com a mínim una vegada durant la cursa.

2. METODOLOGIA

La idea d'aquest treball és poder predir quin es l'ordre final de la cursa. Gràcies a la informació extreta de la web oficial de la Formula 1, amb la posterior construcció i la transformació de la base de dades a un format Excel, s'ha tingut accés a unes dades molt complicades d'obtenir. A partir d'aquestes dades i després de realitzar un treball exhaustiu de filtratge i de depuració, s'han anat estudiant els temps que es consideren útils per predir el resultat final.

L'obtenció i la construcció de dades va patir problemes, ja que a finals de l'any 2017 va canviar la propietat de la Formula 1. Ara a mans de una empresa anomenada *Liberty Media*. Aquest canvi de direcció va provocar una variació en el format de la pàgina web principal de la competició i la desaparició de la majoria de les dades necessàries per fer l'estudi. Aquest contratemps va dificultar molt la construcció de la base de dades en la part inicial del treball.

Si observem l'estructura de les dades amb les que es treballa tenim onze variables en total però si descartem dues (*minuts* i *segons*) que hem utilitzat per a la construcció d'una d'elles *time*, a més de la variable *Number* que tampoc serà necessària, ja que es tracta del dorsal del pilot i per tant, és un identificador, ens quedem amb vuit variables.

A la base de dades, tenim una mateixa variable registrada en diferents moments al llarg del cap de setmana del Gran Premi de Catalunya, això es coneix com a dades longitudinals. En concret, tenim els temps realitzats per cada pilot en les diferents sessions de pràctiques. Per tant, al tenir moltes dades per a cada individu ens trobem en una situació en la que es fa necessària la utilització de models mixtes per el seu anàlisi. Aquest tipus de models no s'han estudiat en profunditat en el grau, el que implica passar per un procés d'autoaprenentatge.

2.1 INTRODUCCIÓ

El principal objectiu a l'hora de crear models estadístics és poder provar les hipòtesis d'investigació eliminant el biaix de les estimacions i tenint en compte la variabilitat que presenten les dades. En aquest context, la modelització estadística ens ajuda a predir i avaluar resultats, i poder observar l'associació entre les respectives variables explicatives i la variable resposta.

La problemàtica més gran a l'hora de realitzar l'anàlisi de la variància és que el nombre d'observacions per individu no és idèntic i les dades estan recollides en intervals de temps irregulars. És per aquest motiu, (no tenir un disseny balancejat) que no podem aplicar

l'anàlisi de la variància clàssic; ja que, tot i que pot contemplar mesures repetides, necessita que el nombre d'observacions per element sigui idèntic i simultani.

Per poder pal·liar aquest inconvenient, podem utilitzar els models mixtes, aquests esdevenen importants quan les dades a tractar estan agrupades. La premissa d'independència entre casos és necessària pels models lineals generalitzats, però el fet que els grups compostos per les dades tinguin relació (un mateix grup té més d'una observació, per exemple), fa que s'hagi d'afegir un component aleatori al model.

Així doncs, com que els casos no es poden considerar independents, s'ha d'ampliar el model lineal generalitzat per potenciar-lo mitjançant els models lineals mixtes.

Hi ha diferents situacions que poden donar pas a dades agrupades:

- Dissenys de blocs aleatoritzats, on els diferents tractaments d'interès s'apliquen a cada bloc.
- Dissenys de mesures repetides, on cada unitat experimental contribueix amb més d'una observació, però no existeix estructura temporal ni espacial.
- Dissenys jeràrquics o multinivell, on hi ha diverses capes d'agrupament.
- Dissenys de dades longitudinals, on existeix una covariable temporal que suposa una estructura particular de dependència entre observacions.

2.2 FORMULACIÓ

L'especificació del model amb dos efectes aleatoris creuats per a l' i -èssim i j -èssim grup es defineix com:

$$Y_{ij} = X_{ij}\beta + Z_i b_i + Z_j b_j + \varepsilon_{ij}$$

- β : Vector d'efectes fixes (paràmetres) de llargada p .
- b_i, b_j : Efectes aleatoris (variables aleatòries).
- X_i i Z_i, Z_j : matrius de disseny del efectes fixos i aleatoris (dim. $n_i \times p$ i $n_i \times q, n_j \times q$ respectivament).
- ε_{ij} : Error aleatori (provinent de les variables aleatòries).
- G : Matriu de covariàncies de l'efecte aleatori (dim. $q \times q$).
- R : Matriu de covariàncies dels errors.

El LMM és un model condicional representat com $Y_i|u \sim N(X_i\beta + Z_iu, R)$ on s'assumeix que $u \sim N(0, G)$. Els paràmetres relacionats en la modelització del valor esperat correspon als efectes fixos, i els paràmetres relacionats en la modelització de l'estructura i propietats de la variància correspondria als efectes aleatoris.

La distribució marginal d' Y_i ve donada per $Y_i \sim N(X_i\beta, V = Z_i^T G Z_i + R)$ on la matriu $V_i = Z_i^T G Z_i + R$ descriu la variància marginal de la resposta.

Per tant els respectius residus es calculen com:

$$\begin{aligned} e_i^{(cond)} &= Y - X\beta - Zb_i \\ \text{on } e_i^{(cond)} &\sim N(0, \sigma^2) \text{ i } b_i \sim N(0, \sigma^2 D) \\ e_i^{(marg)} &= Y - X\beta \end{aligned}$$

En el cas dels models mixtes, la població objecte de la inferència estadística és la població d'on provenen les seves unitats experimentals. D'aquesta manera, en un cas concret, la unitat experimental d'un individu pot donar resultats extrapolables a la població d'individus.

En un cas amb dades correlacionades, les unitats es troben agrupades en clústers (internament homogenis), on cada clúster representa un nivell del factor aleatori. Conèixer una part de les dades d'un clúster permetrà fer prediccions sobre noves unitats del mateix clúster, entre les quals existirà certa correlació, doncs un factor aleatori genera una estructura de correlacions.

Per a l'estimació en els models mixtes cal tenir en compte que els efectes aleatoris no són observats directament, ja que l'interès es centra en els paràmetres fixes. Gràcies a això, es pot concentrar la versemblança global sobre aquests paràmetres d'interès.

Llavors, quan les dades són normals, la solució consisteix en utilitzar la versemblança marginal o integrada, que es pot avaluar de forma exacta. Quan, per contra, les dades no tenen normalitat, cal aproximar la integral:

L'objectiu és estimar l'alfa i la beta a partir de la funció de versemblança marginal:

$$\begin{aligned} L(\beta, \alpha|y) &= f_y(y|\beta, \alpha) = \int_b f_{Y|b}(y|b, \beta, \alpha) f_b(b|\alpha) db \\ I_{ML} &= -\frac{1}{2} \log|V(\alpha)| - \frac{n}{2} \log(2\pi) = -\frac{1}{2} (Y - X\hat{\beta}_{ML}(\alpha))' V^{-1}(\alpha) (Y - X\hat{\beta}_{ML}(\alpha)) \end{aligned}$$

Aquesta estimació ML per a α està esbiaixada al no contemplar la pèrdua d'informació (graus de llibertat) que s'atribueix al fet d'estimar la β . Es corregeix doncs amb la màxima versemblança restringida (REML).

$$\begin{aligned} I_{REML} &= -\frac{1}{2} \log |V(\alpha)| - \frac{n-p}{2} \log(2\pi) - \frac{1}{2} \log |X'V(\alpha)X| \\ &= -\frac{1}{2} (Y - X\hat{\beta}_{ML}(\alpha))' V^{-1}(\alpha) (Y - X\hat{\beta}_{ML}(\alpha)) \end{aligned}$$

Després d'obtenir la $\hat{V}(\hat{\alpha})$, ja es poden estimar definitivament α i β . Per màxima versemblança:

$$\begin{aligned} \hat{\beta}_{ML} &= (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y \\ b_{ML} &= GZ'\hat{V}^{-1}(Y - X\hat{\beta}) \end{aligned}$$

Si G i R són conegudes, $\hat{\beta}$ és el millor estimador lineal sense biaix de β , i \hat{b} és el millor predictor lineal sense biaix de b .

La inferència per a efectes fixes es realitza mitjançant el Test de Wald, consistent en l'estimació de la matriu de covariàncies del vector $(\hat{\beta} - \beta, \hat{b} - b)$:

$$C = \begin{pmatrix} X'\hat{R}^{-1}X & X'\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}Z + \hat{G}^{-1} \end{pmatrix}^{-1}$$

Aquesta matriu però, sol subestimar la variabilitat mostral de $(\hat{\beta}, \hat{b})$, doncs no incorpora la correcció associada a la incertesa extra provocada per l'estimació dels paràmetres G i R .

Amb això ja es poden dur a terme contrastos d'hipòtesis sobre paràmetres del model:

$$H: L \begin{bmatrix} \beta \\ b \end{bmatrix} = 0$$

En cas que L consisteixi en una sola filera (un contrast és per a un únic paràmetre o només té un grau de llibertat):

$$t = L \begin{bmatrix} \beta \\ b \end{bmatrix} / \sqrt{L\hat{C}L'} \approx t_v$$

En el supòsit que L tingui un rang superior a 1, es té la següent expressió:

$$F = \frac{\begin{bmatrix} \hat{\beta} \\ \hat{b} \end{bmatrix}' L' (L\hat{C}L')^{-1} L \begin{bmatrix} \hat{\beta} \\ \hat{b} \end{bmatrix}}{Rang(L\hat{C}L')}$$

De forma anàloga, l'estadístic segueix, aproximadament, una distribució F amb $Rang(L\hat{C}L')$ graus de llibertat al denominador i que han de ser estimats per al denominador.

Una tècnica alternativa per a estimar els efectes fixes pot ésser el test de raó de versemblança (LR test), l'estadístic del qual és el següent:

$$-2\log\lambda_N = -2\log \frac{L_{ML}(Model\ 1)}{L_{ML}(Model\ 0)}$$

On el model 1 és un cas particular del model 0. Aquest test permet comparar models amb estructures d'efectes fixes niades i és vàlid per a models ajustats només per ML, no per REML.

Per a les components de la variància en LR test, l'interès es centra en garantir la inferència sobre β , on la correcta modelització d' α permet afinar els mètodes inferencials i permet interpretar la variabilitat.

Cal tenir en compte que una sobreparametrització d' α provocarà una estimació ineficient mentre que, per contra, la infraparametrització provocaria una inferència inadequada. α ha de contenir els elements que han participat al disseny, i és recomanable explorar estructures amb sentit real.

Per a comparar models niats es pot procedir de forma similar a la que s'ha vist en el cas dels efectes fixes. Canvia l'estructura de les covariàncies però s'utilitza el mateix model d'efectes fixes:

$$-2\log\lambda_N = -2\log \frac{L_{REML}(Model\ 1)}{L_{REML}(Model\ 0)}$$

Que és vàlid per a models ajustats mitjançant ML i REML.

A partir de les proves LR es poden comparar models mixtes, que tindran versemblances majors com més alt sigui el nombre de paràmetres dins el model. En cas de comparar models que no siguin niats, cal penalitzar en funció, també, del nombre de paràmetres utilitzats. En aquest punt, es fan servir habitualment mesures d'adequació a les dades com poden ser:

$$AIC = -2\log L(\text{Model 1}) + 2p$$
$$BIC = -2\log L(\text{Model 1}) + \log(n)p$$

on p representa el número de paràmetres del model i n és el nombre total d'observacions.

La validació de les suposicions del model hauria de considerar totes les premisses que han intervingut en la seva construcció, tals com la correcta especificació del model fixe i de l'estructura de variàncies. L'error del model i els efectes aleatoris (per separat) segueixen una distribució Normal amb estructura també especificada.

Per últim, el diagnòstic del model es pot representar mitjançant els següents diagrames:

- Histograma de residus condicionals.
- Histograma dels efectes aleatoris.
- Diagrama de dispersió dels residus marginals envers valors predits.
- Diagrama de dispersió dels residus condicionals envers valors predits.
- Punts influents.

El software R, i el paquet lme4 entre altres, seran les eines utilitzades per computar els estimadors del model lineal mixt:

- **lmer**: Estimació
- **fixef**: β (efectes fix)
- **ranef**: b (efectes aleatoris)
- **resid**: e_i (residus)
- **summary**: Wald per β
- **anova**: LRT

2.3 VALIDACIÓ DEL MODEL

Per la validació del model lineal mixt, es pot fer a partir d'anàlisis de gràfiques. Es pot realitzar un gràfic de residus estandarditzats contra els valors ajustats per tal d'observar que la variància és constant. També plantejar un gràfic de normalitat dels residus i dels efectes aleatoris, per comprovar si les dades es distribueixen com una normal. D'altra banda, també seria convenient realitzar un gràfic dels valors predits contra les observacions del model, per tal d'observar la precisió del model.

2.4 PLOTLY

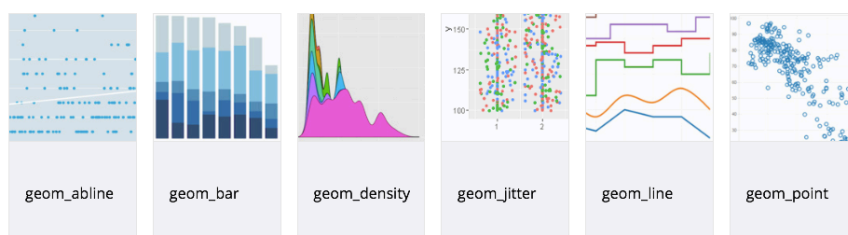
Últimament es fa molta èmfasi en cuidar la presentació en el moment de mostrar els resultats finals d'un estudi. Aconseguir transmetre la informació d'una forma clara i entenedora és un dels objectius principals de qualsevol projecte. És per això que he decidit contribuir en aquest procés d'innovació utilitzant una sèrie de animacions gràfiques a l'hora de exposar els resultats.

Les funcions gràfiques utilitzades del treball són: `barplot`, `hist`, `plot`, `dotplot`, `scatterplot`, `xyplot`, `boxplot`, `qqnorm`, `formattable` i `plot_ly`.

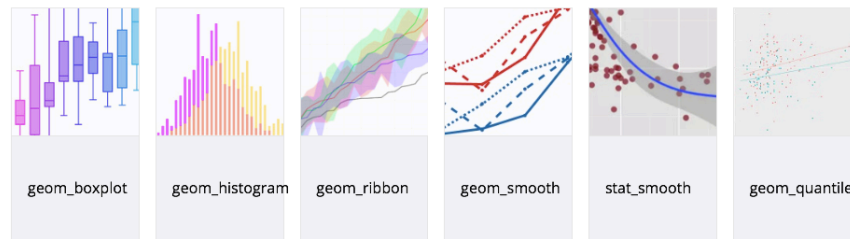
El *package* més important utilitzat per presentar alguns resultats de l'estudi es "Plotly". Aquest paquet a partir de la funció `plot_ly`, permet la construcció de gràfics interactius amb animacions a més dels estàtics convencionals. Una de les moltes avantatges és que no només ofereix una fotografia d'un moment determinat sinó que pots recórrer tot l'àmbit temporal i seleccionar el que creguis més oportú.

A continuació es presentaran alguns exemples del diferents gràfics que es poden construir amb aquest *package*.

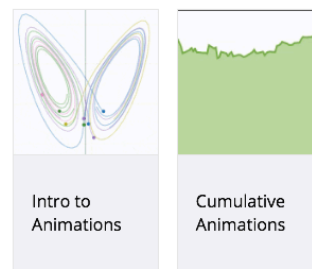
- Gràfics basics:



- Gràfics estadístics:



- Gràfics animats:

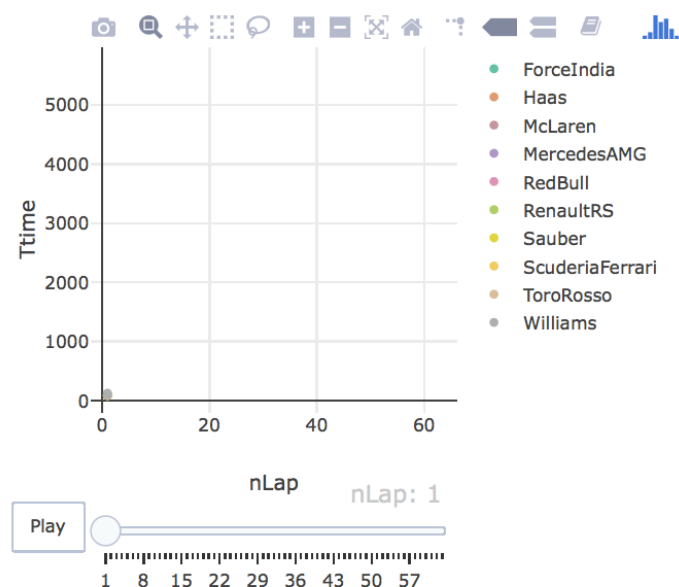


Per més informació, l'adreça electrònica d'aquest paquet és: <https://plot.ly/r/>

A part de l'exemple animat que es mostra a continuació, també s'ha utilitzat aquest paquet per il·lustrar el rendiment i la degradació dels diferents pneumàtics pels Gran Premis de Catalunya dels anys 2016 i 2017. Figura 14: Rendiment i degradació segons compost i any.

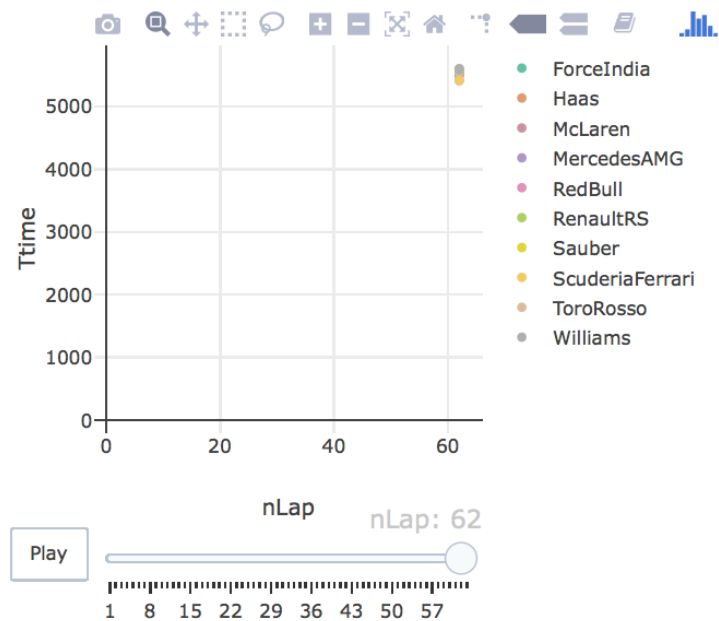
En aquest cas, també s'ha utilitzat la part de gràfics animats per representar la evolució dels temps de tots els pilots al llarg de la cursa. El resultat és el següent:

FIGURA 1: SITUACIÓ INICIAL DE LA CURSA



Si passem al final de la cursa clicant el botó (Play) que trobem a la part inferior a l'esquerra, es pot observar com els punts que representen els pilots, van incrementant temps total de cursa, a mesura que passen les voltes. El gràfic resultant és (Figura 2: Situació final de la cursa):

FIGURA 2: SITUACIÓ FINAL DE LA CURSA



Com es fa complicada la diferenciació dels punts, donat per l'amplitud dels eixos, "Plotly" ofereix una funció que busca automàticament les observacions dins del gràfic (Autoescale).

FIGURA 3: "AUTOESCALE" DE LA SITUACIÓ FINAL DE LA CURSA



Ara ja es distingeixen millor els pilots (Figura 3: "Autoescale" de la situació final de la cursa). En el cas de voler saber exactament quin pilot representa cada punt, només has de posar el cursor de l'ordinador a sobre del punt escollit.

2.5 FORMATTABLE

Aquest paquet està dissenyat per aplicar en vectors i data frames un format de visualització fet a mida per les dades i així facilitar la presentació de dades i per incloure més informació.

En general, aquesta funció ajuda a d'identificar i/o magnificar els valors de la taula ràpidament. Aquesta eina és senzilla d'utilitzar i a part de les principals funcions que s'explicaran a continuació, el paquet disposa d'una ampla varietat d'eines visuals.

Una de les possibles aplicacions és que podem fer, per exemple, és que els valors positius siguin de color verd, els negatius en vermell i els zeros negres mitjançant la creació d'una funció que realitzi una transformació condicional des del valor al codi HTML (`change`). Un altre exemple també podria ser crear una altra funció que faci que els valors superiors a la mitjana es mostrin en negreta deixant els altres sense canvis(`prev_score`, `cur_score`).

id	prev_score	cur_score	change
1	10	8	-2
2	8	9	1
3	6	7	1
4	8	8	0
5	8	9	1

L'ús de funcions integrades com `color_tile ()` i `color_bar ()` fa que sigui més fàcil comparar la magnitud de les columnes especificades (`rating`, `market_share`). Amb aquesta funció la cel·la de cada xifra de la columna del data frame s'omple de forma proporcional al valor de la cel·la en qüestió amb el color desitjat.

De vegades, també és útil aplicar una funció a una àrea perquè totes les cel·les de la zona comparteixin un punt de referència (`price`).

id	price	rating	market_share	revenue	profit
1	10	5	10.00%	55,000.00	25,300.00
2	15	4	12.00%	36,400.00	11,500.00
3	12	4	5.00%	12,000.00	(8,200.00)
4	8	3	3.00%	(25,000.00)	(46,000.00)
5	9	4	14.00%	98,100.00	65,000.00

Per més informació sobre la funció `formattable` es pot trobar a la direcció web descrita a continuació:

<https://cran.r-project.org/web/packages/formattable/formattable.pdf>

En el nostre cas es va construir Figura 17: Diferències entre la volta de classificació i la volta més ràpida de les pràctiques, la per veure la diferència de temps entre la millor volta de les sessions de pràctiques i la classificació.

3. CONSTRUCCIÓ DE LA BASE DE DADES

Aquest treball se centra en l'estudi de les voltes realitzades en les sessions d'entrenaments lliures, en la modalitat automobilística de la Fórmula 1. L'objectiu és observar i analitzar la metodologia de treball dels equips i veure si realment es pot predir el resultat final de la cursa amb les dades amb les quals treballen els tècnics dels equips.

Per poder començar a assolir l'objectiu, es van extreure dades i informació de curses d'anys anteriors de les següents direccions electròniques:

- Gran Premi de Catalunya 2017: <https://www.fia.com/events/fia-formula-one-world-championship/season-2017/eventtiming-information-21>
- Gran Premi de Catalunya 2016: <https://www.fia.com/events/fia-formula-one-world-championship/season-2016/event-timing-information-13>

En aquest treball s'han utilitzat les dades del Gran Premi de Catalunya 2017. Les dades del 2016 són d'ajuda per comprovar si realment el codi que utilitzen és vàlid per a altres situacions. De cada adreça electrònica anterior, es pot descarregar tota la informació necessària que hi ha disponible del Gran Premi de Fórmula 1. Malgrat tot, l'únic obstacle és el format en el qual estan penjades aquestes dades (PDF), que provoca que s'hagi de modificar-les manualment i convertir-les al programa informàtic Excel per tal de poder tractar-les. Es van anar seleccionant només les dades extretes de cada volta neta, és a dir, no es una volta ni de sortida ni de entrada a boxes. A més, també va ser necessari cercar informació complementària que no s'especificava en el fitxer disponible.

He agafat d'exemple per explicar el procés de construcció de la base de dades del treball les voltes de la FP3¹⁰ del pilot Sebastian Vettel que porta el dorsal número 5 i condueix per l'equip ScuderiaFerrari (Imatge 4: Voltes de S.Vettel a FP3).

IMATGE 4: VOLTES DE S.VETTEL A FP3

5 S. VETTEL

NO	TIME	NO	TIME
1 P	11:03:06	5	19:51.170
2	10:56.459	6	1:20.456
3	1:23.406	7 P	1:56.103
4 P	1:52.429		

¹⁰ FP3, sigles de *free practice 3*, fa referència als tercers entrenaments lliures.

En el cas anterior (tal com mostra la imatge), la sessió de pràctiques va començar a les 11 hores del matí. El primer temps de tots indica que el cotxe va sortir del box a les 11 hores, 3 minuts i 6 segons. Per saber si una volta pertany a l'entrada al box, el temps està marcat amb una "P" just al costat del número de la volta. El temps que s'indica després d'una volta d'entrada a boxes són els minuts que triga el pilot a tornar a passar per la línia de sortida, és a dir, el temps que està dins el box sumat al temps que triga a donar una volta al circuit.

En aquest cas, el temps que triga en tornar a passar per la línia de sortida són 10 minuts, 56 segons i 459 mil·lèsimes de segon. La volta següent, indicada amb un temps de 1:23.406, ja és una volta neta. Després, el pilot torna a entrar a boxes (P) i, en tornar a creuar la línia de sortida, triga gairebé 20 minuts. La sessió acaba amb una altra volta neta més. En total, el pilot ha fet dues voltes que podrien ser objecte d'estudi en la sessió de pràctiques FP3(Figura 4: Dades en format Excel).

FIGURA 4: DADES EN FORMAT EXCEL

Number	Driver	Team	Engine	Time	Session
	5 S.Vettel	ScuderiaFerrari	Ferrari	1:23.406	FP3
	5 S.Vettel	ScuderiaFerrari	Ferrari	1:20.456	FP3

Una altra dificultat, va ser que tampoc hi havia les dades del compost amb què es va fer cada volta. Per sort, en finalitzar cada sessió de proves, sempre apareix un resum de la sessió (Imatge 5: Resum de la FP3) en la qual s'indica els tipus de pneumàtics utilitzats.

IMATGE 5: RESUM DE LA FP3

F1® THIRD PRACTICE CLASSIFICATION			CATALUNYA		#SpanishGP	
			LAP TIME	GAP	LAPS	TYRE STRATEGY
1	Kimi Raikkonen	Ferrari	1:20.214		20	H M S S S S
2	Sebastian Vettel	Ferrari	1:20.456	+0.242	7	H M S
3	Lewis Hamilton	Mercedes	1:20.555	+0.341	12	S S S S S S
4	Valtteri Bottas	Mercedes	1:20.868	+0.654	7	S
5	Max Verstappen	Red Bull Racing	1:21.025	+0.811	14	H M M S
6	Daniel Ricciardo	Red Bull Racing	1:21.249	+1.035	15	H M M S
7	Nico Hulkenberg	Renault	1:21.670	+1.456	13	M S S
8	Felipe Massa	Williams	1:21.746	+1.532	19	H M M S S S
9	Carlos Sainz	Toro Rosso	1:21.835	+1.621	18	H H H M M S S
10	Fernando Alonso	McLaren	1:22.093	+1.879	16	H M M S
11	Romain Grosjean	Haas	1:22.128	+1.914	21	H H H M S
12	Kevin Magnussen	Haas	1:22.214	+2.000	12	H M S
13	Sergio Perez	Force India	1:22.237	+2.023	19	M M M S
14	Esteban Ocon	Force India	1:22.297	+2.083	22	M M M S
15	Daniil Kvyat	Toro Rosso	1:22.391	+2.177	15	H M M S
16	Marcus Ericsson	Sauber	1:22.513	+2.299	20	H H H M M S
17	Lance Stroll	Williams	1:22.574	+2.360	19	H S S S S S
18	Jolyon Palmer	Renault	1:22.755	+2.541	13	M S
19	Stoffel Vandoorne	McLaren	1:22.853	+2.639	15	M M S
20	Pascal Wehrlein	Sauber	1:22.974	+2.760	19	H H H M M S

Com s'observa a la llista anterior, el pilot seleccionat com a exemple (Sebastian Vettel) va utilitzar primer un pneumàtic dur usat, marcat a la dreta amb una H sobre fons negre. Ara bé, no va fer cap volta neta amb aquest compost, ja que va sortir i va entrar a boxes a la mateixa volta. Aquesta volta té el nom de "volta d'instal·lació" i molts equips la realitzen. Aquests equips utilitzen la primera volta de la sessió per comprovar, amb els sensors del cotxe, que tot estigui en bon estat i funcioni correctament. Després de parar en boxes, van canviar els pneumàtics del cotxe i van utilitzar un compost mitjà nou (marcat amb una M), i, finalment, després d'entrar una altra vegada al box, van utilitzar un pneumàtic tou nou (marcat amb una S) per acabar la sessió (Figura 5: Dades en format Excel amb el compost utilitzat en cada volta).

FIGURA 5: DADES EN FORMAT EXCEL AMB EL COMPOST UTILITZAT EN CADA VOLTA

Number	Driver	Team	Engine	Tyre	Time	Session
5	S.Vettel	ScuderiaFerrari	Ferrari	M	1:23.406	FP3
5	S.Vettel	ScuderiaFerrari	Ferrari	S	1:20.456	FP3

Finalment, es va haver de passar la variable *Time* (temps expressat en minuts, segons i mil·lèsimes de segon) a una sola magnitud per poder treballar amb ella més fàcilment a l'R. La magnitud escollida van ser els segons. Per tant, una volta amb un temps d'1:27.715 passa a ser una volta de 87.715 segons. També es va crear una altra variable (*nLap*) que indicava el número de la volta de cada pilot en cada sessió de pràctiques (Figura 6: Base de dades definitiva).

FIGURA 6: BASE DE DADES DEFINITIVA

Number	Driver	Team	Engine	Tyre	Time	Session	minuts	segons	time	nLap
5	S.Vettel	ScuderiaFerrari	Ferrari	M	1:23.406	FP3	1	23,406	83,406	1
5	S.Vettel	ScuderiaFerrari	Ferrari	S	1:20.456	FP3	1	20,456	80,456	2

Així doncs, la base de dades en format Excel queda amb aquestes variables:

- **Number:** dorsal del pilot, un identificador. Un total de 20 dorsals.
- **Driver:** Inicial del nom i cognom del pilot. Un total de 20 pilots.
- **Team:** Equip per el qual condueix el pilot. Un total de 10 equips.
- **Engine:** Marca del motor que porta el monoplaça. Un total de 4 motoristes.
- **Tyre:** Tipus de compost amb el qual s'ha fet la volta. Un total de 3 compostos.
- **Time:** Temps en realitzar la volta.
- **Session:** Sessió de pràctiques en què s'ha fet la volta. Un total de 3 sessions.
- **minuts:** Minuts de la volta.
- **segons:** Segons que han passat des del minut anterior.

- **time:** Temps de la volta senyalitzat en segons.
- **nLap:** Número de volta.

A més també disposem d'una sèrie de bases de dades complementaries extretes de la mateixa direcció electrònica que la base de dades de les sessions de pràctiques mencionada prèviament:

- Base de dades de la velocitat màxima (en kilòmetres per hora) establerta en la volta més ràpida de les tres sessions per cada pilot.
- Base de dades dels temps dels *pit stop* realitzats a la cursa.
- Base de dades amb 64 observacions (temps de cada volta) per a cada pilot que va finalitzar la cursa.
- Base de dades amb el millor temps de cada pilot a la classificació.

4. LES DADES

Un cop obtingudes les dades, es duu a terme la seva anàlisi descriptiva per fer un primera aproximació visual i comprovar si es poden formular algunes hipòtesis sobre els resultats i possibles conclusions que es poden arribar a extreure.

Per començar l'anàlisi descriptiva; s'indica, on hi ha la descriptiva univariada, el nombre de voltes per cada variable d'estudi escollida. El primer que sobta és la presència de tres pilots de l'equip RenaultRS. Aquests pilots són: S.Sirotkin, J.Palmer i N.Hulkenberg. Les normes de la Fórmula 1 són molt clares i esclareixen que només dos pilots de cada escuderia poden participar en la cursa. Per tant, un d'ells quedarà descartat.

Si s'analitzen les sessions, en les quals van participar els tres pilots amb les voltes netes completades entre parèntesis, s'obté la taula següent:

TAULA 4: PILOTS TITULARS DE L'EQUIP RENAULTRS

SESSIÓ	S.Sirotkin	J.Palmer	N.Hulkenberg
FP1	*(7)		*(14)
FP2		*(33)	*(30)
FP3		*(9)	*(7)
TOTAL	7 voltes	42 voltes	51 voltes

Segons la Taula 4: Pilots titulars de l'equip RenaultRs, tot sembla indicar que S.Sirotkin no participarà en la cursa. A fi de poder assegurar aquesta conjectura, a l'adreça electrònica adjuntada a continuació confirma que els pilots titulars són N.Hulkenberg i J.Palmer:

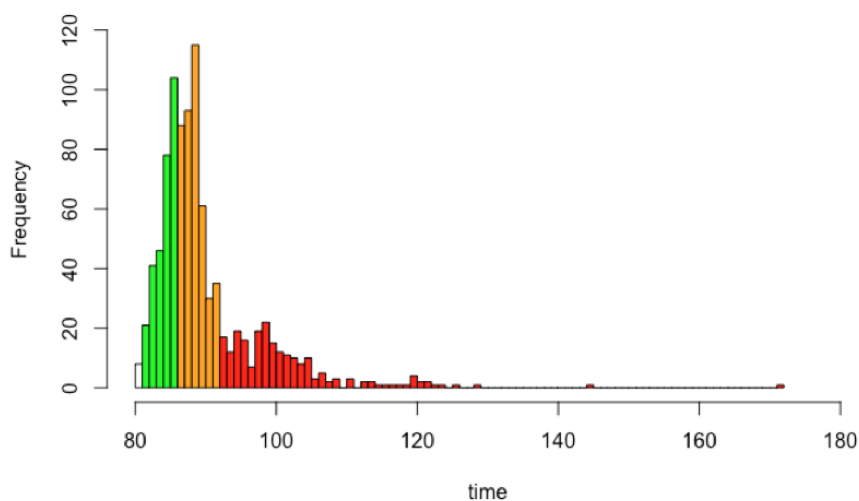
- https://es.wikipedia.org/wiki/Temporada_2017_de_Fórmula_1

És rellevant destacar el gran nombre de voltes que ha fet l'equip Williams amb els seus dos pilots F.Massa i L.Stroll, que situen a l'equip com el que més voltes netes va completar al llarg de totes les sessions de pràctiques. També cal recalcar les poques voltes realitzades pel pilot F.Alonso. Això és degut al fet que no va donar cap volta a la sessió FP1, segurament a causa d'algun problema mecànic que li va impedir participar. Pel que fa a la quantitat de voltes realitzades per cada motorista, s'observa clarament la falta de registres amb el motor Honda. Aquest fet era d'esperar, ja que Honda només subministra motors a l'equip McLaren, mentre que la resta en subministra a tres equips diferents.

Respecte el pneumàtic més utilitzat pels equips participants ha estat el mitjà (M) seguit del tou (S). La sessió de proves en les quals es donen més voltes és la FP2 i en la que menys la FP3, ja que aquesta última és la més curta de les tres.

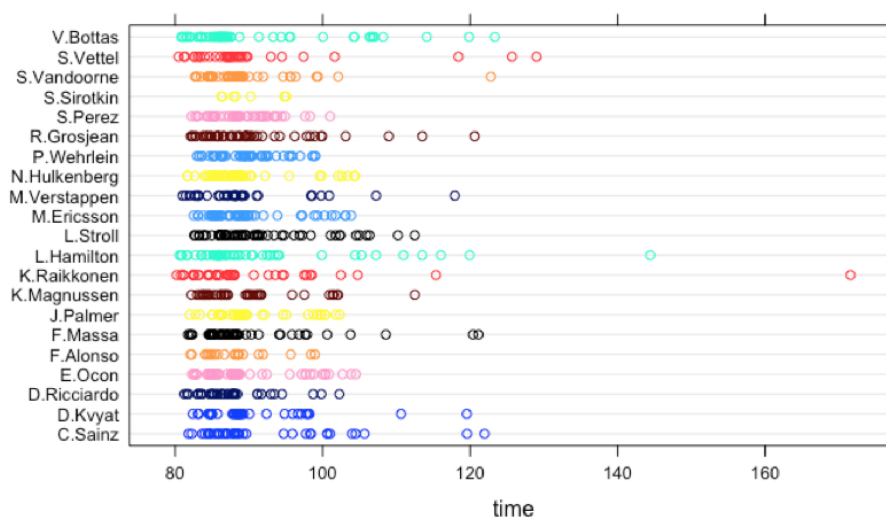
Quant a la Figura 7: Freqüència dels temps, tenim representades les freqüències dels temps. Per a aquest gràfic, es té en compte un criteri propi en el qual s'utilitzen el primer i tercer quartil per categoritzar les voltes. Considerem com a una volta bona (color verd) si el seu temps és inferior als 85,32 segons, com a normal (color taronja) si el seu temps està entre l'interval de 85,33 i els 91,50 segons i, finalment, com a dolenta (color vermell) si triga més de 91,51 segons.

FIGURA 7: FREQUÈNCIA DELS TEMPS



Fixant-se en la freqüència del temps estratificada pel pilot (Figura 8: Freqüència dels temps per pilot), resulta que tots fan uns temps semblants, exceptuant els pilots de ScuderiaFerrari, MercedesAMG i RedBull, que sembla que són els que protagonitzen els temps més ràpids. Així doncs, es pot plantejar una primera hipòtesi dels equips més competitius. Aquesta idea es reforça si s'analitza el gràfic de la freqüència dels temps per cada equip.

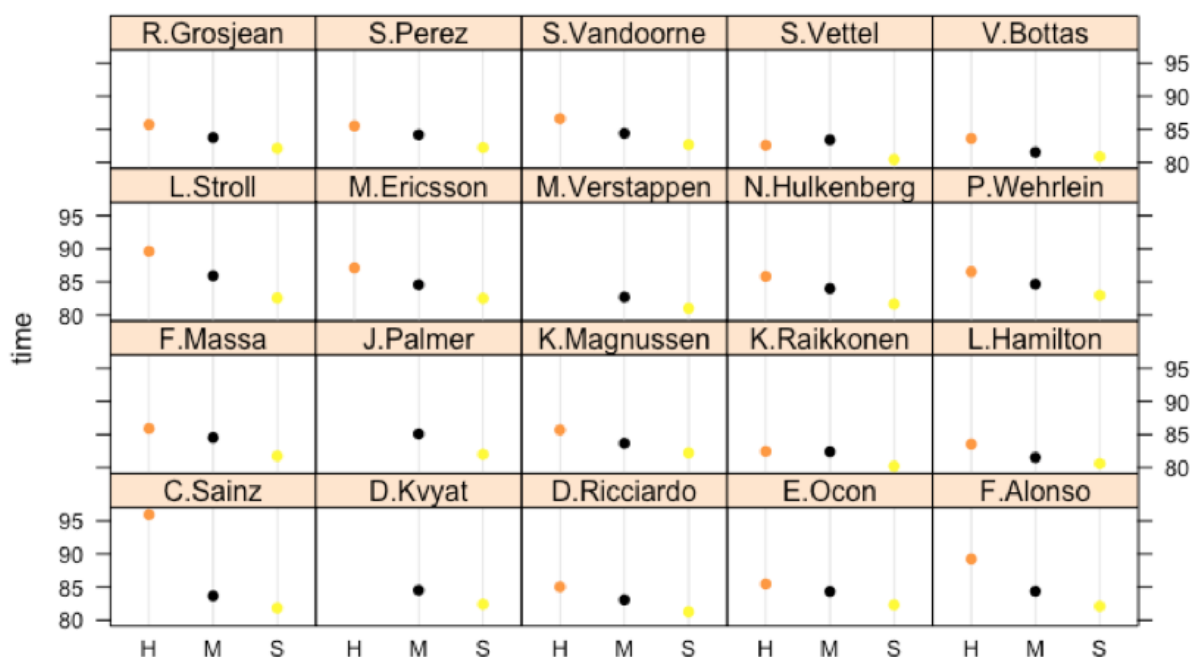
FIGURA 8: FREQUÈNCIA DELS TEMPS PER PILOT



4.1 SIMULACIÓ DE CLASSIFICACIÓ

A l'hora de participar en una cursa, tenir un bon cotxe i un bon pilot són requisits imprescindibles si vols ser competitiu. Tot i això, hi ha una altra variable determinant en el cas que la cursa estigui igualada: l'estratègia de pneumàtics. Saber quin moment és l'oportú per parar a boxes i fer el canvi de rodes adequadament, pot fer variar la posició final dels pilots de forma radical. És per això que s'ha d'analitzar el rendiment de tots els compostos que es té a disposició. En la Figura 9: Voltes ràpides per compost i pilot, es poden distingir els millors temps de cada pilot realitzats amb cada compost.

FIGURA 9: VOLTES RÀPIDES PER COMPOST I PILOT

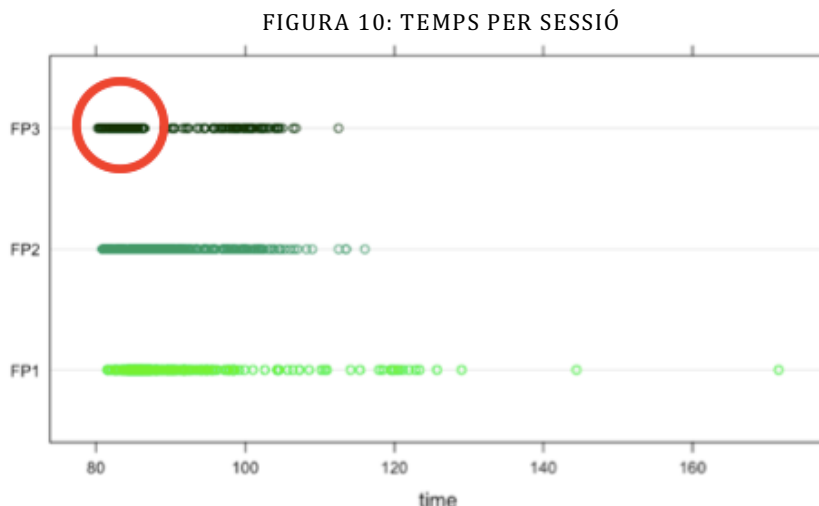


Un cop observada la taula anterior, es contempla que el pneumàtic S és el més ràpid i que el H és el més lent, tot i que s'ha d'analitzar individualment el cas de cada pilot. Es podria donar el cas que un monoplaça aconseguís fer treballar millor un compost més dur, és a dir, que degradés tant el tou que al final de la volta cronometrada no tingués suficient goma al pneumàtic per acabar fent la volta més ràpida. Tot i que hi ha pilots que no han utilitzat els pneumàtics H, la tònica general fa indicar que com més tou sigui el compost, més ràpida serà la volta. Hi trobem excepcions, com en el cas de S. Vettel i de K. Raikkonen, pilots de Scuderia Ferrari, que han marcat un temps més ràpid amb el compost H que amb el compost M.

El que sembla una mica clar, tot i que s'ha de confirmar amb un test, és que el pneumàtic S és el més adequat per fer la volta ràpida. Per tant, sembla que serà aquest el que triïn en el moment de fer la classificació. Aleshores, es planteja un test de l'Anova amb "Tyre" com a

única variable explicativa del model. S'utilitzaran les dades d'una base de dades elaborada a partir dels millors temps de cada pilot amb cada pneumàtic. El resultat obtingut presenta diferències significatives entre compostos, sent el tou (S) i el mitjà (M) gairebé quatre i dos segons, respectivament, més ràpid per volta que el compost dur (H).

Si es miren els temps per volta fets a cada sessió de proves (Figura 10: Temps per sessió), es pot veure que els millors temps són cada vegada més ràpids. No obstant això, el que crida l'atenció és un núvol de punts de voltes ràpides a la FP3 (senyalitzat amb un cercle vermell al gràfic següent):



Concretament, es veu com hi ha un salt de rendiment entre aquesta sèrie de voltes i la resta que hi trobem dins de la mateixa sessió. Aquest fet, que tampoc es troba a cap altra sessió, porta a plantejar-se la hipòtesi que potser la simulació de la classificació es fa en la tercera sessió de pràctiques, just abans de la classificació real.

FIGURA 11: MILLOR TEMPS DE LES FP PER CADA PILOT

GridPosition	Number	Driver	Team	Engine	Tyre	Time	Session	time
1	7	K.Raikkonen	ScuderiaFerrari	Ferrari	S	1:20.214	FP3	80.214
2	5	S.Vettel	ScuderiaFerrari	Ferrari	S	1:20.456	FP3	80.456
3	44	L.Hamilton	MercedesAMG	Mercedes	S	1:20.595	FP3	80.595
4	77	V.Bottas	MercedesAMG	Mercedes	S	1:20.868	FP3	80.868
5	33	M.Verstappen	RedBull	Renault	S	1:21.025	FP3	81.025
6	3	D.Ricciardo	RedBull	Renault	S	1:21.249	FP3	81.249
7	27	N.Hulkenberg	RenaultRS	Renault	S	1:21.67	FP3	81.670
8	19	F.Massa	Williams	Mercedes	S	1:21.746	FP3	81.746
9	55	C.Sainz	ToroRosso	Renault	S	1:21.835	FP3	81.835
10	30	J.Palmer	RenaultRS	Renault	S	1:21.992	FP2	81.992
11	14	F.Alonso	McLaren	Honda	S	1:22.093	FP3	82.093
12	8	R.Grosjean	Haas	Ferrari	S	1:22.128	FP3	82.128
13	20	K.Magnussen	Haas	Ferrari	S	1:22.214	FP3	82.214
14	11	S.Perez	ForceIndia	Mercedes	S	1:22.237	FP3	82.237
15	31	E.Ocon	ForceIndia	Mercedes	S	1:22.297	FP3	82.297
16	26	D.Kvyat	ToroRosso	Renault	S	1:22.391	FP3	82.391
17	9	M.Ericsson	Sauber	Ferrari	S	1:22.513	FP3	82.513
18	18	L.Stroll	Williams	Mercedes	S	1:22.574	FP3	82.574
19	2	S.Vandoorne	McLaren	Honda	S	1:22.693	FP2	82.693
20	94	P.Wehrlein	Sauber	Ferrari	S	1:22.974	FP3	82.974

Tot seguit a la Figura 11: Millor temps de les FP per cada pilot, podem veure que tots els pilots van fer la volta més ràpida de les tres sessions a la FP3 amb el compost S. Tots a excepció de dos: J.Palmer i S.Vandoorne, que la van fer també amb el compost tou però a la FP2.

Si analitzem l'ordre de les voltes més ràpides realitzades per cada pilot es pot veure com les sis primeres posicions estan ocupades per només tres equips (ScuderiaFerrari, MercedesAMG i RedBull). La diferència entre el primer i el sisè és d'aproximadament un segon. Si a partir de la posició número set mirem la quantitat de pilots que hi ha en el següent segon, estan dins gairebé tots a excepció de P.Wehle (1.3 segons). Per tant, tot fa indicar que tenim dos grups en quant a rendiment: Els tres equips mencionats prèviament i la resta. El gran repte que ens planteja aquesta igualtat dins de cada grup, és la dificultat d'ordenar correctament la graella, ja que perdre mig segon en la volta decisiva de classificació implica sortir molt més enrere.

A l'hora de realitzar un test de l'Anova per comprovar si existeixen diferències significatives entre els millors temps segons cada equip, s'obté un resultat no significatiu. Per tant, amb aquest model amb "Team" com a única variable explicativa, es pot concloure que el fet de conduir per a un equip de bo no garanteix assolir un bon temps al resultat. Aquest resultat pot arribar a ser sorprenent, ja que tots hem escoltat alguna vegada la frase: "Si a la Fórmula 1 no tens un bon cotxe, no guanyaràs mai". A l'adreça electrònica adjuntada a continuació podrem llegir un article d'un diari esportiu en el que un pilot parla justament d'aquest tema.

- <http://www.marca.com/claromx/motor/formula1/2018/05/05/5aed476f22601d655d8b4597.html>

Si s'analitzen les diferències de l'equip de referència (ForceIndia) sobre la resta, s'observa que hi ha tres escuderies que destaquen de manera positiva com hem vist prèviament. Aquestes són: ScuderiaFerrari, MercedesAMG i RedBull. Totes tres són de mitja, com a mínim, un segon per volta més ràpid que l'equip de referència. El problema està en que hi ha dos grups molt heterogenis entre ells però homogenis dins d'ells. Aleshores, si la diferència és tan gran, el dubte recau en per què el model no troba cap diferència entre ells. El problema recau en aquest fet.

Així mateix, es podria plantejar un model semblant només amb la variable "Engine". La dificultat que es presenta és que el motor no és l'únic factor que pot explicar el temps per volta; sinó que es coneixen altres aspectes com les suspensions, l'aerodinàmica i l'electrònica, entre d'altres, que també influeixen. L'únic pas més o menys raonable seria comparar els temps dels equips oficials, és a dir, els que construeixen els seus propis motors. Aquest serien: Ferrari amb ScuderiaFerrari, Mercedes amb MercedesAMG, Renault amb RenaultRS i, finalment, Honda amb McLaren. Si es descarten els equips "client", que paguen als equips oficials per tenir un motor, les diferències entre els motors passen a ser

significatives. El model no troba diferències entre el motor Ferrari i Mercedes. Tot i això, sí troba diferències entre aquest primer i l'Honda i Renault, sent dos i un segon i mig més lents per volta, respectivament. Aquest resultat té credibilitat en el supòsit que les diferències de *xassis*¹¹ o suspensions fossin nul·les.

Altrament, si s'utilitzen les dades de les velocitats puntes de cada pilot en la seva volta més ràpida dels entrenaments, es pot observar que per tots els equips, a excepció dels que utilitzen motor Mercedes, tenir més velocitat en la recta implica aconseguir millors temps. Una alta velocitat punta es pot aconseguir de dues maneres: per tenir un motor bo amb molts cavalls de potència o pel fet de reduir molt el *drag*¹² del teu monoplaça, deixant els *flaps* dels alerons en una posició més horitzontal. Això pot fer que es plantegi alguna hipòtesi: potser els equips amb motor Mercedes (sabent que tenen un motor amb molts cavalls de potència) han sacrificat una mica la velocitat punta per igualar-se amb la resta, incrementant la incidència dels *flaps* per poder anar més ràpid als revolts. Aquest fet també ajuda a que hi hagi menys degradació en els pneumàtics, ja que el cotxe relisca menys en prendre un revolt o en el moment de la tracció al sortir d'aquest.

IMATGE 6: COMPARACIÓ ALERÓ POSTERIOR AMB POC I MOLT DRAG



Font: <https://www.motorsport.com/f1/photo/main-gallery/red-bull-rb12-rear-wing-sochi-8469293/>

A la imatge de l'esquerra hi ha un exemple d'un aleró del darrere del cotxe amb una càrrega aerodinàmica baixa. A la dreta hi ha un *flap* en una posició més vertical, això provoca més resistència aerodinàmica a les rectes, amb la conseqüent pèrdua de velocitat punta, però amb un millor pas pel revolt.

Després d'aquest parèntesis en el que hem explicat les diferències entre els millors temps segons l'equip i motor, tornem a la hipòtesi de l'existència d'una simulació de classificació.

¹¹ *Xassis*, consisteix en una estructura interna que sosté, aporta rigidesa i dona forma a un vehicle.

¹² *Drag*, de l'anglès, es tradueix com a arrossegar. És una paraula que s'utilitza habitualment per referir-se a la resistència aerodinàmica que ofereix el monoplaça. Un cotxe amb molt de *drag* ofereix molta resistència aerodinàmica.

En resum, sembla que aquesta simulació es fa a l'última sessió d'entrenaments lliures (FP3) amb el pneumàtic més ràpid (S) que disposen els equips.

Fins ara s'ha identificat la simulació de classificació, en la qual els temps més ràpids de les tres sessions es feien als entrenaments dels dissabtes previs a la classificació. Per tant, l'ordre esperat de la graella de sortida es troba a la Figura 11: Millor temps de les FP per cada pilot. Tanmateix, encara queda trobar la simulació de cursa.

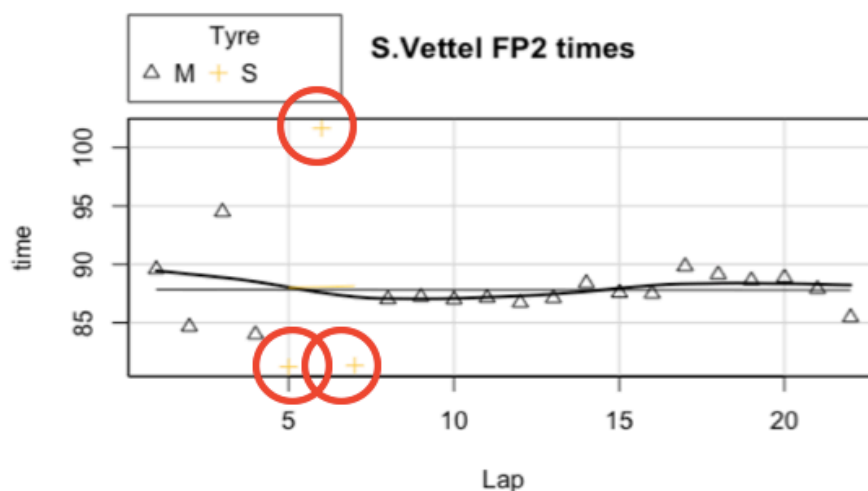
4.2 SIMULACIÓ DE CURSA

Per tal d'identificar una simulació de cursa, es necessitaria cercar una sèrie de voltes consecutives en les quals els temps siguin, de forma gradual, cada vegada més lents. Aquesta pèrdua de temps d'una volta respecte a la següent és deguda a la degradació dels pneumàtics. Si bé, també es podria produir l'efecte contrari en el cas que la simulació de cursa es faci amb un pneumàtic amb molt poca degradació i, per tant, l'efecte de la reducció de pes del monoplaça, provocat pel consum de combustible, pot ser més gran que el de la degradació, fent que els temps siguin millors. El problema d'utilitzar un compost amb aquestes característiques és que els temps per volta poden ser molt més lents que amb un pneumàtic més tou.

Intuint que el compost tou és gairebé quatre segons més ràpid per volta i que una parada en boxes dura aproximadament vint-i-dos segons, si aquest compost té una durada de més de cinc voltes, ja seria més efectiu que el dur. Tot això suposant que la degradació fos igual en els dos compostos.

Si es pren com a exemple el pilot de ScuderiaFerrari, Sebastian Vettel, i s'analitzen les voltes que va fer a la segona sessió de lliures, s'obté la Figura 12: Temps de S.Vettel a la FP2 per compost:

FIGURA 12: TEMPS DE S.VETTEL A LA FP2 PER COMPOST



En total, aquest pilot va fer 22 voltes durant la FP2. A l'inici de la sessió realitza quatre voltes amb el compost mitjà (marcat amb una M i un triangle negre), seguides de tres voltes amb el compost més tou (marcat amb una S i un signe sumant groc) dels que disposen els equips. Des del primer moment sembla que s'estableix un patró que es va repetint entre la primera volta i la setena, independentment del canvi de pneumàtics. Sembla que els temps són contradictoris, ja que fa una volta ràpida seguida d'una lenta. Curiosament, els temps més ràpids es produeixen en aquestes set voltes tan diferents, concretament en el moment que porta els pneumàtics S (indicades en el gràfic anterior amb un cercle vermell).

La resposta que es dona a aquest comportament és provocada pel tipus de motor dels monoplaques. L'any 2014, el reglament de la Fórmula 1 va fer un gran canvi en termes del tipus de propulsor que havien de subministrar els motoristes als equips. Aquest havien de ser motors híbrids de sis cilindres amb turbo, tot al contrari dels de vuit cilindres atmosfèrics que hi havia fins aquell moment. Les bateries que introduïen els nous propulsors havien de treballar conjuntament amb el motor tèrmic i, així, transmetre més potència a les rodes del darrere. L'única dificultat era que aquestes bateries només es podrien carregar amb l'energia produïda a les frenades. És aleshores quan entra en joc l'estratègia d'equip. En el moment de voler fer la volta ràpida, el que interessa és que les bateries estiguin totalment carregades per disposar de la màxima potència durant una única volta.

Últimament, tots els equips han anat perfeccionant l'eficiència dels propulsors fins al punt de tenir diferents mapes de motors i així modificar el comportament del cotxe per després seleccionar el que més convingui en un moment determinat. En la gràfica anterior es poden identificar tres tipus de mapes de motor:

- Classificació: màxima potència en una sola volta combinant el motor tèrmic i elèctric sense recuperació de energia a les frenades. (Voltes ràpides)
- Recàrrega de la bateria: el cotxe només funciona amb el motor tèrmic mentre es carreguen les bateries a les frenades. (Voltes lentes)
- Cursa: potència combinada del motor tèrmic i el elèctric amb recuperació d'energia a les frenades. (Voltes constants)

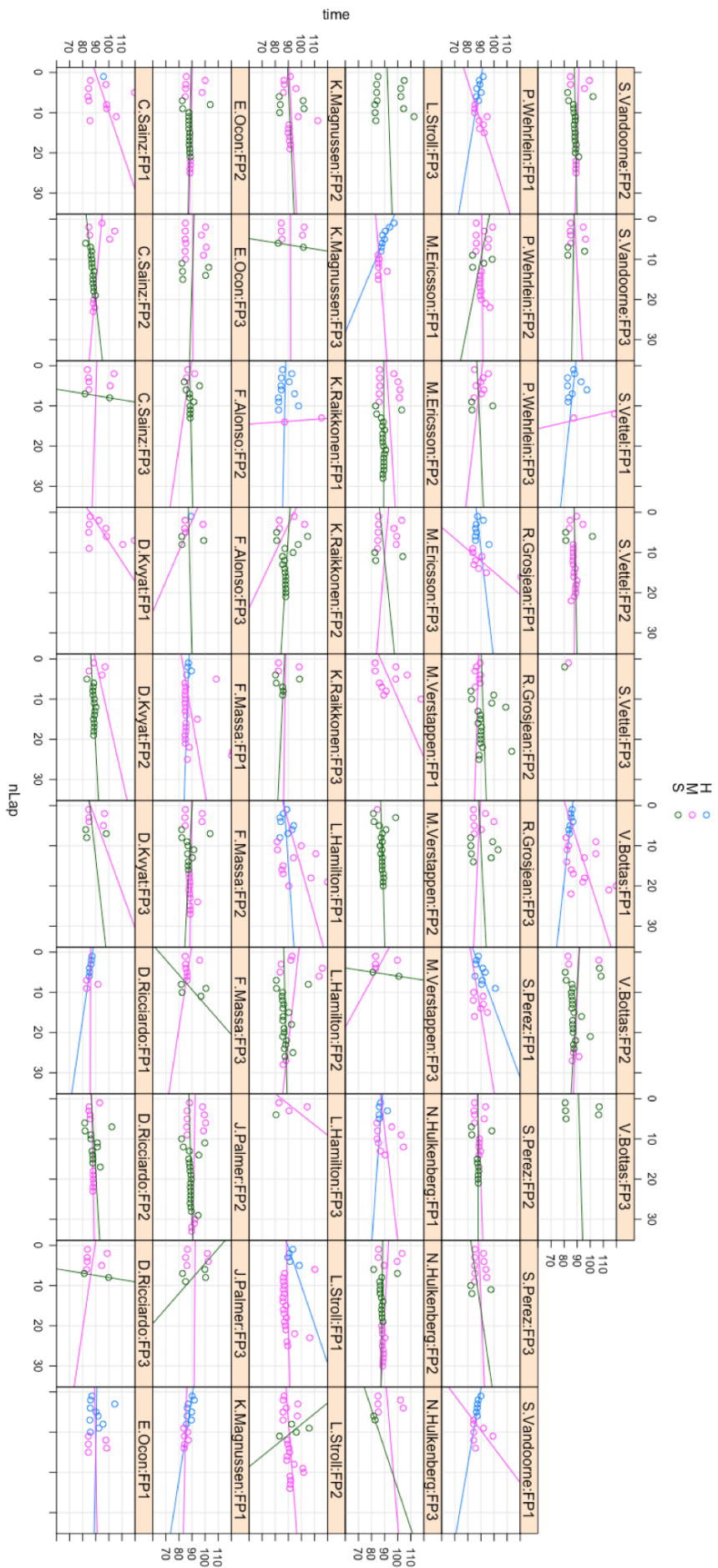
És per això que es troba una diferència de temps tan gran en les voltes més ràpides de la sessió i la resta. La primera volta amb el compost S és amb màxima potència, ja que ha tingut abans la volta de sortida de boxes per poder recarregar les bateries. Aquesta volta no està representada al gràfic, ja que només hi són les voltes netes. A la segona volta canvia el mapa motor a recàrrega de bateria i, a la tercera volta, ja amb les bateries completament carregades, passa una altra vegada al mapa de classificació. Finalment, acaba amb el que sembla una simulació de cursa, ja que les voltes són seguides, realitzades amb el mateix

compost i no hi ha una gran variabilitat entre elles. Tot i això, aquesta hipòtesi s'haurà de confirmar posteriorment.

A continuació adjunto una direcció web que parla exactament del fet descobert de la funció dels mapes de motor en un monoplaça de Formula 1.

- <http://soymotor.com/videos/24241-como-funcionan-los-mapas-motor-de-mclaren-honda-keny500-simracing>

Si es representen totes les voltes de tots els pilots que participaran en el Gran Premi de Fórmula 1 al circuit de Montmeló, surt un conjunt de gràfics en els quals es comencen a distingir alguns patrons. A les gràfiques que analitzen el comportament dels residus es troben moltes observacions extremes, fet que provoca que no es comportin com una Normal, el que impedeix validar el model resultant de les dades. Per tal de reafirmar el que diuen les gràfiques anteriors, es crea un BoxPlot amb les observacions de cada compost per veure les dades atípiques. La majoria de punts que queden fora del rang interquartílic, són voltes fetes amb els compostos més utilitzats (tou i mitjà. Aquest fet segurament és degut al canvi del mapa del motor per recarregar les bateries, com s'ha comentat prèviament. Per tant, aquestes voltes de recàrrega no són representatives del ritme real del monoplaça en una situació de cursa.



Tot seguit, es fa una primera depuració, donat el gran soroll que presenten les dades, aplicant una sèrie de condicions per ser capaços d'identificar les voltes que podrien formar part d'una simulació de cursa. Els criteris principals que s'han de complir per ser considerades "voltes de simulació de cursa" són les següents:

- Els temps s'han hagut de fer amb el mateix compost, amb un mínim de cinc voltes i de forma consecutiva. Es considera que cinc voltes són les necessàries per visualitzar el nivell i l'evolució de la degradació dels pneumàtics en el monoplaça.
- S'ha establert un ràtio per determinar si una volta és suficientment semblant a l'anterior, per tal de ser capaços de descartar les simulacions de classificació o voltes amb errors de pilotatge. Per tant, si aquest ràtio és més gran que l'establert, aquesta volta quedarà descartada i es tornarà a inicialitzar el procés d'identificació de la simulació de cursa.

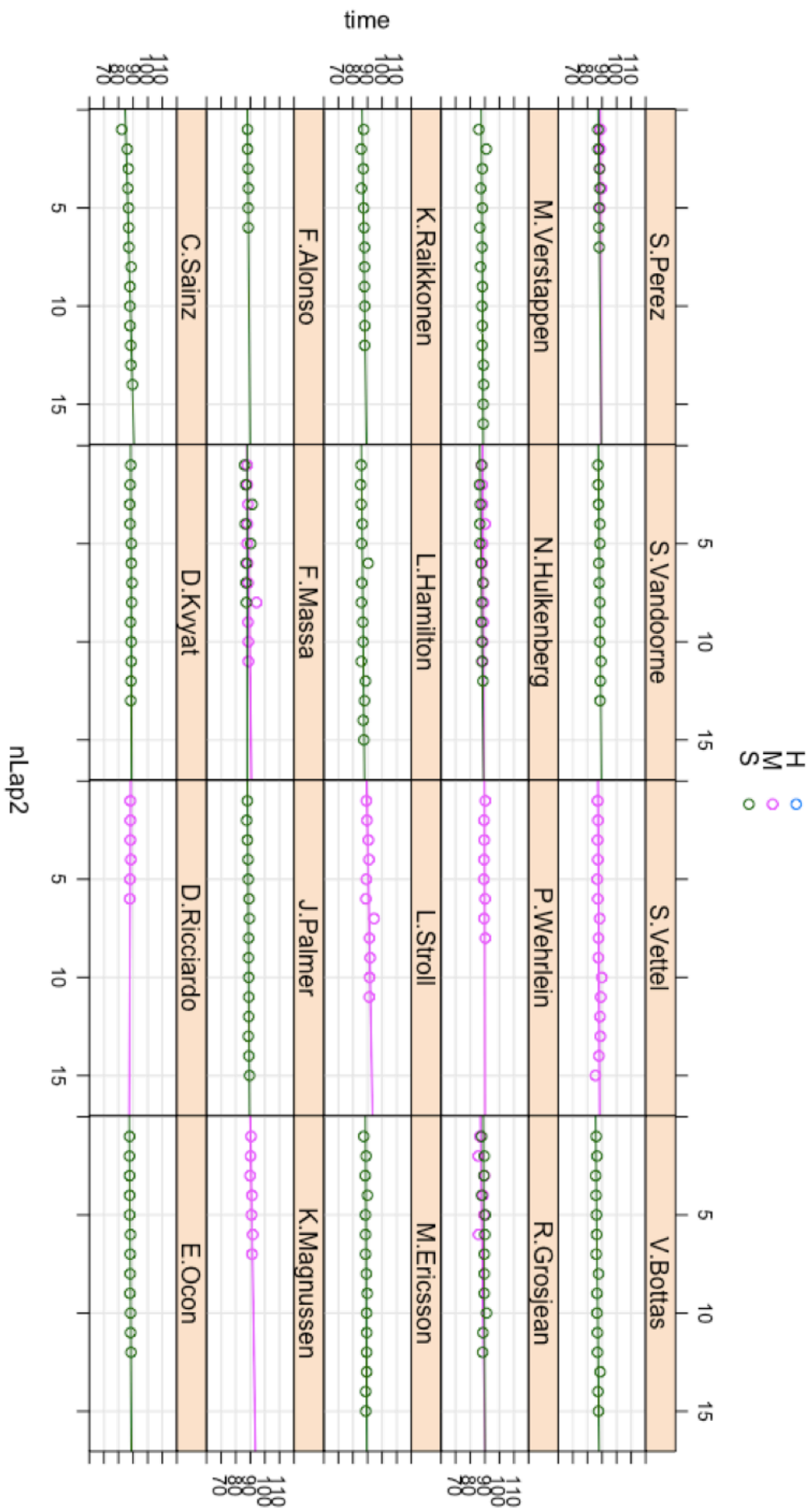
Una vegada aplicats aquests criteris, s'aconsegueix com a mínim una simulació de cursa per a cada pilot i totes realitzades a la FP2. Hi ha set pilots que també en fan a la FP1; però s'han fet amb els compostos més durs que, com s'ha vist, és el menys competitiu dels tres. Això porta a determinar que al igual que la simulació de classificació es feia a la FP3, la simulació de cursa la fan a la FP2.

Ara que ja tenim identificades les simulacions de cursa realitzades per cada pilot, es crea una nova variable anomenada "nLap2" que numera les voltes de cada simulació, inicialitzant-se a 1 quan es canvia el compost al començar una altra sèrie de voltes. Amb aquesta nova variable serem molt més precisos a l'hora de calcular els efectes de les covariables. Sobretot en les relacionades amb els pneumàtics.

Si s'observa el gràfic final, tots els pilots van fer una simulació de cursa segons els criteris marcats i la majoria va escollir el pneumàtic tou per fer-la. Sorprèn el fet que els pilots de ScuderiaFerrari són els únics que van decidir fer les simulacions amb compostos diferents. Personalment, crec és una bona pràctica, ja que així pots comparar millor el rendiment i el nivell de degradació tant del tou com del mitjà en un mateix cotxe.

Tot i que s'ha assolit una bona primera depuració que ha permès extreure algunes conclusions, encara hi ha alguna volta que sobresurt per sobre la resta. Per tant, es treuen tretze temps que es considerava que provocaven residus alts.

Finalment, es comprova com els valors dels residus són acceptables i aquests es distribueixen com una Normal. En darrer terme, després de fer una exhaustiva neteja d'algunes observacions innecessàries, s'ha aconseguit identificar les simulacions de cursa realitzades en les sessions de pràctiques de tots els pilots.



A més, aquest criteri per aïllar la simulació de cursa s'ha provat amb les dades del GP de 2016. El resultat obtingut també va ser satisfactori. A partir d'aquí, ja es pot plantejar el model predictiu que ajudarà a vaticinar l'evolució dels temps dels pilots al llarg de les voltes en el Gran Premi de Montmeló.

Recordant el que s'ha mencionat, si es planteja el resum del model (Figura 13: Resultat del Model (m3)) en el que tenim en compte les variables *Team* i *Driver* com a efectes aleatoris, es troben alguns efectes significatius en les variables explicatives restants:

FIGURA 13: RESULTAT DEL MODEL (M3)

```
> summary(m3)
Linear mixed model fit by REML t-tests use Satterthwaite
approximations to degrees of freedom [lmerMod]
Formula: time ~ nLap2 + Tyre + (1 | Team) + (1 | Driver)
Data: dadesFin

REML criterion at convergence: 503.9

Scaled residuals:
    Min       1Q   Median       3Q      Max
-2.67584 -0.55416 -0.06523  0.47521  2.90157

Random effects:
 Groups   Name      Variance Std.Dev.
Driver   (Intercept) 0.4275   0.6538
Team     (Intercept) 0.5840   0.7642
Residual                0.3191   0.5649
Number of obs: 252, groups: Driver, 20; Team, 10

Fixed effects:
              Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  88.263402   0.305556 11.770000 288.862 < 2e-16 ***
nLap2         0.122296   0.009849 233.180000 12.417 < 2e-16 ***
TyreS        -1.026723   0.140429 233.800000 -7.311 4.18e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A continuació s'exposen les interpretacions resultants dels efectes del model per a cada covariable:

- *nLap2*: mostra que l'efecte del temps que es guanya pel fet de reduir el pes del monoplaça a causa del consum de combustible, ja que cada volta que passa el cotxe pesa menys, no es suficient per evitar una pèrdua 0,122 dècimes per volta a causa de l'existència de degradació en els pneumàtics.
- *TyreS*: mostra que l'efecte de canviar de pneumàtic mitjà a tou et fa ser un segon més ràpid cada volta.

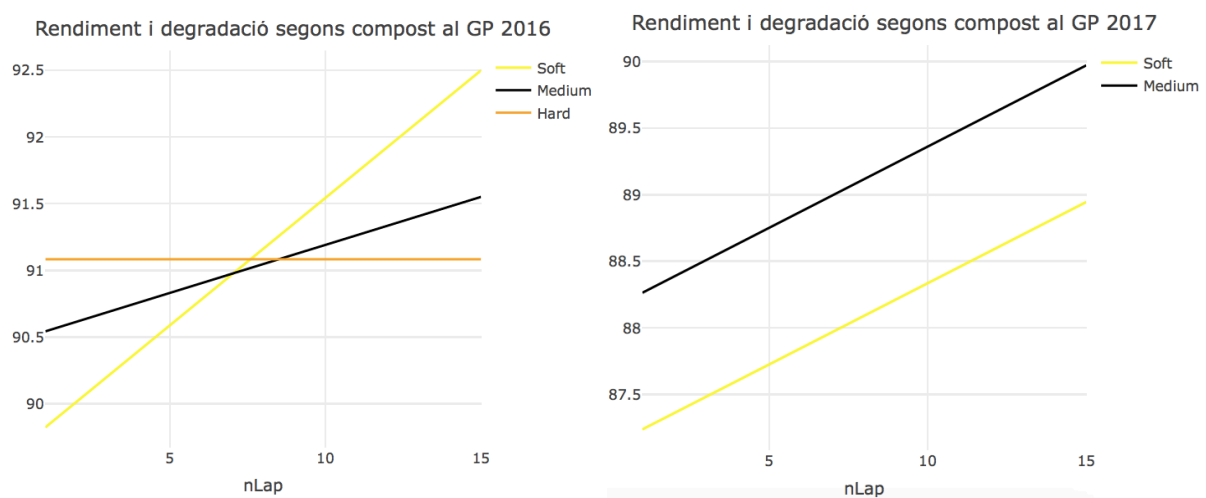
També es va analitzar la interacció entre la variable *TyreS* i *nLap2* en un altre model per veure si la degradació era diferent entre dos compostos i aquesta va sortir no significativa. Per tant es degraden per igual.

En aquest cas com que els dos compostos presenten el mateix nivell de degradació, segurament els equips preferirien que els seus pilots estiguessin el màxim de voltes amb el pneumàtic tou. Però en algun moment de la cursa tots els pilots hauran de canviar de compost, ja que estan obligats a utilitzar dos compostos diferents durant el transcurs d'aquesta.

També s'ha aplicat aquesta metodologia a les dades del Gran Premi del 2016, obtenint, en aquest cas, un resultat que permet caracteritzar millor la degradació i el rendiment dels pneumàtics.

A continuació, i de manera puntual, es mostren les diferències de rendiment i de degradació dels compostos per l'any 2016 i 2017 en el Gran Premi de Catalunya, Montmeló:

FIGURA 14: RENDIMENT I DEGRADACIÓ SEGONS COMPOST I ANY



Es pot observar com en el cas de l'any 2017 hi ha un salt de rendiment significatiu entre compostos, igual que en l'any 2016. Tot i això, la degradació en aquest últim ja no es comporta de la mateixa manera. En el cas del 2016, el compost més ràpid es degrada més i el compost més dur no presenta cap desgast. Per tant, el tou és més ràpid que el mitjà i el dur fins a la setena o vuitena volta. A partir d'aquí, hi ha un punt d'inflexió en el qual el dur passa a ser el més competitiu dels tres.

Finalment, el model resultant del GP de l'any 2017 segons el model obtingut és:

$$Y_{ij} = 88.263 + 0.122 * nLap2 - 1.026 * TyreS + b_i + b_j + \varepsilon_{ij}$$

$i = (1, \dots, 20) = Team$

$b_i \sim N(0, 0.654) Team$

$j = (1, 2) = Driver$

$b_j \sim N(0, 0.764) Driver in Team$

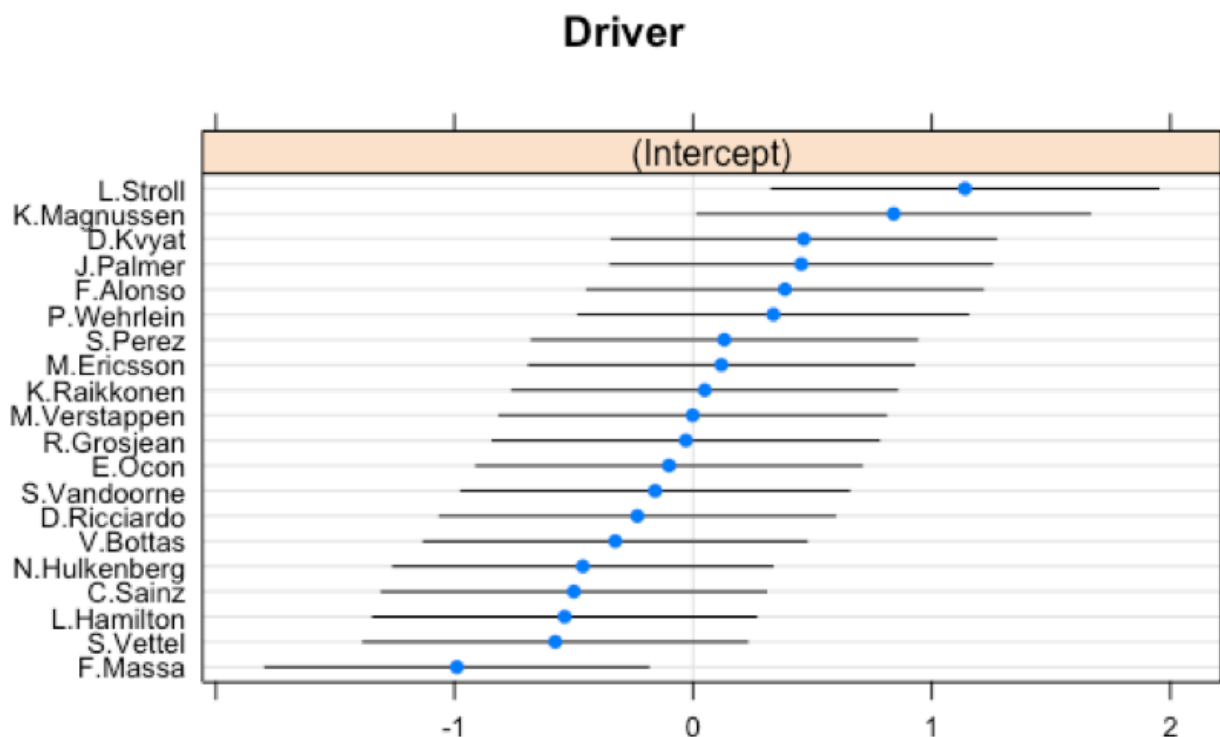
$Y_{ij} : Temps de la volta del Driver j del Team i$

$\varepsilon_{ij} \sim N(0, 0.565) Residual$

Una vegada es disposa del model resultant de les dades es passa a part d'estratègia de parades. Gràcies a la informació de cada compost i després plantejar un problema de programació lineal en el que es vol minimitzar el temps total de la cursa, s'arriba a la conclusió que s'han de fer trenta-una voltes amb el M (mitjà) i trenta-cinc amb el S (tou). Tot i que a partir de la volta deu es podrien plantejar els equips fer un *undercut*.

Per tal d'acabar aquesta part d'identificació i interpretació de la simulació de cursa, també és rellevant mirar si hi ha diferències entre pilots dins del mateix equip (Figura 15: Efecte aleatori dels pilots).

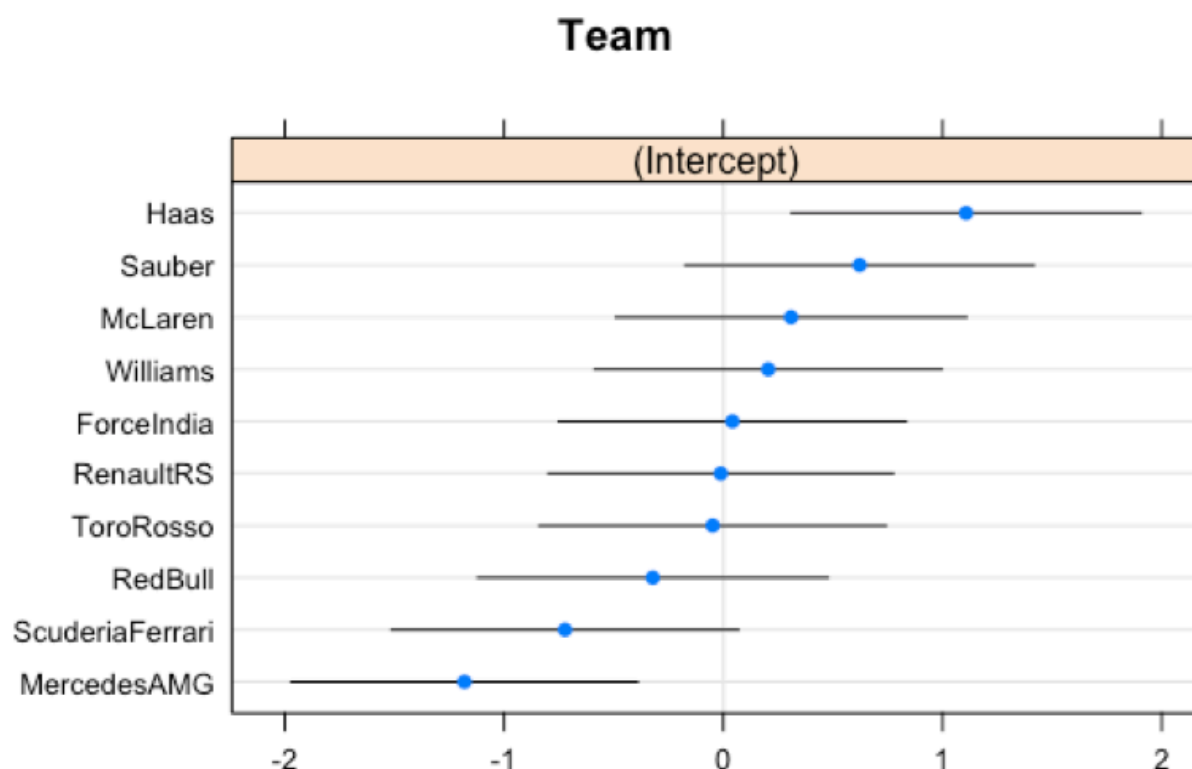
FIGURA 15: EFECTE ALEATORI DELS PILOTS



Un cop analitzat el gràfic anterior, cal destacar que l'única parella de pilots en la qual es troben diferències significatives és a l'equip Williams (F.Massa i L.Stroll). Per tant, la diferència de rendiment ve provocada pels pilots. Cal dir que aquesta va ser la primera temporada de L.Stroll, pilot de divuit anys i el més jove de la graella; mentre que el seu company d'equip F.Massa és un pilot que porta tretze anys participant en aquesta competició. Per tant, l'experiència en la Fórmula 1 podria ser un factor que influeixi en el rendiment. El fet de no trobar diferències significatives entre la resta de parelles de pilots podria donar-se al fet que els test de pretemporada es fan des de fa molts anys en aquest circuit, i també que aquest traçat s'utilitza en categories inferiors a la Formula 1, així que tots els pilots el coneixen a la perfecció.

Quant a la diferència de rendiment dels monoplaces (Figura 16: Efecte aleatori dels equips), es troben diferències significatives entre MercedesAMG amb Haas i Sauber, i ScuderiaFerrari amb només Haas. De tota manera, es torna a confirmar que destaquen positivament els equips ScuderiaFerrari i MercedesAMG sobre la resta, quedant com a favorits per guanyar el Gran Premi de Catalunya, sobretot aquest últim.

FIGURA 16: EFECTE ALEATORI DELS EQUIPS



4.3 EVOLUCIÓ DE LA CURSA

Per poder saber si el ritme de cursa estimat és realment representatiu del real, s'ha dividit la cursa en tres parts. Les voltes escollides per a la comparació són: volta 2, i no la volta 1, ja que la sortida provoca una certa incertesa i per tant, poden produir-se canvis de posició a més de possibles incidents. També s'ha de tenir en compte que els pilots comencen la cursa des de la graella i no amb la inèrcia que duen en la simulació. La següent volta és la 10, ja que s'ha recorregut la distància suficient per poder observar si la degradació és semblant a l'estimada, si el ritme de cursa s'ajusta correctament i perquè no hi ha cap parada a boxes que faci desvirtuar el temps global de les primeres deu voltes. Finalment, l'última volta d'estudi és la 64, i no la 66 (voltes totals del Gran Premi de Catalunya), ja que hi ha una sèrie de pilots que són avançats pel líder de la cursa i per tant, no arriben a completar la distància total del Gran Premi.

TAULA 5: ESTIMACIÓ DE LA EVOLUCIÓ DELS TEMPS EN LA CURSA PER A LES VOLTES 2,10 I 64

Predicció	Volta		
Pilot	2	10	64
L.Hamilton	171.4025	861.9041	5634.169
V.Bottas	171.8267	864.0255	5647.533
S.Vettel	172.2416	866.0996	5660.600
F.Massa	173.2703	871.2432	5693.005
K.Raikkonen	173.4936	872.3601	5700.041
D.Ricciardo	173.7298	873.5411	5707.481
C.Sainz	173.7447	873.6155	5707.950
N.Hulkenberg	173.8933	874.3582	5712.629
M.Verstappen	174.1939	875.8613	5722.099
E.Ocon	174.7211	878.4973	5738.706
S.Vandoorne	175.1408	880.5960	5751.927
S.Perez	175.1850	880.8169	5753.319
D.Kvyat	175.6716	883.2499	5768.647
J.Palmer	175.7242	883.5129	5770.304
F.Alonso	176.2290	886.0367	5786.204
M.Ericsson	176.3206	886.4949	5789.091
P.Wehrlein	176.7560	888.6719	5802.806
R.Grosjean	176.9948	889.8660	5810.328
L.Stroll	177.5288	892.5360	5827.150
K.Magnussen	178.7364	898.5737	5865.187

Per tant, el temps total estimat que trigarà L.Hamilton en completar 64 voltes és de 1:33.54.168 hores.

En el cas real la taula anterior passaria a tenir les següents dades:

TAULA 6: EVOLUCIÓ REAL DELS TEMPS EN LA CURSA PER A LES VOLTES 2,10 I 64

Observat	Volta		
Pilot	2	10	64
L.Hamilton	173.328	861.729	5494.986
S.Vettel	172.794	861.542	5501.687
D.Ricciardo	175.615	874.818	5566.463
S.Perez	176.743	881.279	5603.605
E.Ocon	177.804	884.66	5614.034
N.Hulkenberg	178.397	887.202	5635.445
C.Sainz	178.841	890.052	5640.135
P.Wehrlein	180.153	894.308	5637.276
D.Kvyat	196.031	905.158	5647.521
R.Grosjean	179.105	890.976	5654.867
M.Ericsson	178.559	894.671	5661.05
F.Alonso	179.127	891.347	5676.583
F.Massa	213.987	915.287	5666.641
K.Magnussen	179.213	890.038	5689.876
J.Palmer	204.477	916.079	5685.13
L.Stroll	180.3	896.686	5692.614

Per tant, el temps total real que ha trigat L.Hamilton en completar 64 voltes, és de 1:31.34.985 hores, 2 minuts i 19.183 segons menys de l'estimat. Aquesta diferència pot venir perquè durant la cursa va haver-hi una forta competència per la primera posició entre L.Hamilton (MercedesAMG) i S.Vettel (ScuderiaFerrari). Aquest fet podria haver provocat un increment substancial del ritme de cursa per tal d'intentar quedar un per davant de l'altre o per evitar ser avançat.

4.4 COMPARACIÓ DELS RESULTATS AMB LA REALITAT

Ara que ja es disposa d'una sèrie de resultats, tant de classificació com de simulacions de cursa provinents de les sessions de pràctiques prèvies al Gran Premi de Catalunya, es pot comprovar si aquests resultats obtinguts s'ajusten a la realitat.

Tot i això, la predicció que més dificultats comporta és ordenar la graella. Tot seguit s'han obtingut les següents posicions amb els corresponents temps de volta:

TAULA 7: ORDRE DE LA GRAELLA DE SORTIDA REAL RESPECTE ESTIMAT

Position	RealGrid		EstimatedGrid	
1	L.Hamilton	79,149	K.Raikkonen	80,214
2	S.Vettel	79,2	S.Vettel	80,456
3	V.Bottas	79,373	L.Hamilton	80,595
4	K.Raikkonen	79,439	V.Bottas	80,868
5	M.Verstappen	79,706	M.Verstappen	81,025
6	D.Ricciardo	80,175	D.Ricciardo	81,249
7	F.Alonso	81,048	N.Hulkenberg	81,67
8	S.Perez	81,07	F.Massa	81,746
9	F.Massa	81,232	C.Sainz	81,835
10	E.Ocon	81,272	J.Palmer	81,992
11	K.Magnussen	81,329	F.Alonso	82,093
12	C.Sainz	81,371	R.Grosjean	82,128
13	N.Hulkenberg	81,397	K.Magnussen	82,214
14	R.Grosjean	81,517	S.Perez	82,237
15	P.Wehrlein	81,803	E.Ocon	82,297
16	M.Ericsson	82,332	D.Kvyat	82,391
17	J.Palmer	82,401	M.Ericsson	82,513
18	L.Stroll	82,411	L.Stroll	82,574
19	S.Vandoorne	82,532	S.Vandoorne	82,693
20	D.Kvyat	82,746	P.Wehrlein	82,974

Com es pot observar a la Taula 7: Ordre de la graella de sortida real respecte estimat, els temps milloren notablement de la simulació de classificació a la classificació real. Si es crea una taula per magnificar les diferències de temps, s'obté la Figura 17: Diferències entre la volta de classificació i la volta més ràpida de les pràctiques. Aquesta s'ha fet en *RStudio* utilitzant la funció *formattable*, per tal de facilitar la visualització i la interpretació, tot indicant en color verd els temps dels pilots que han rebaixat el millor temps de les pràctiques en la classificació i en vermell els que no ho han fet.

FIGURA 17: DIFERENCIES ENTRE LA VOLTA DE CLASSIFICACIÓ I LA VOLTA MÉS RAPIDA DE LES PRÀCTIQUES

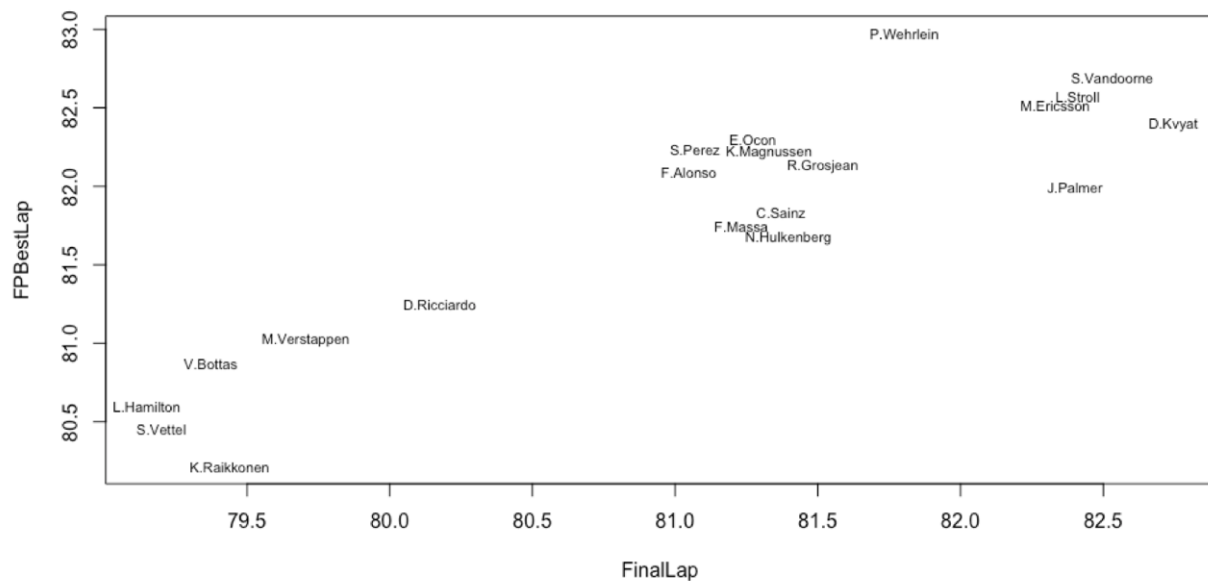
	qfC.Driver	qfC.Team	Dif
3	V.Bottas	MercedesAMG	↓ -1.849
1	L.Hamilton	MercedesAMG	↓ -1.794
5	M.Verstappen	RedBull	↓ -1.628
2	S.Vettel	ScuderiaFerrari	↓ -1.561
8	S.Perez	ForceIndia	↓ -1.419
15	P.Wehrlein	Sauber	↓ -1.411
6	D.Ricciardo	RedBull	↓ -1.322
7	F.Alonso	McLaren	↓ -1.273
10	E.Ocon	ForceIndia	↓ -1.245
11	K.Magnussen	Haas	↓ -1.076
4	K.Raikkonen	ScuderiaFerrari	↓ -0.966
14	R.Grosjean	Haas	↓ -0.744
9	F.Massa	Williams	↓ -0.629
12	C.Sainz	ToroRosso	↓ -0.567
13	N.Hulkenberg	RenaultRS	↓ -0.334
16	M.Ericsson	Sauber	↓ -0.219
18	L.Stroll	Williams	↓ -0.197
19	S.Vandoorne	McLaren	↓ -0.195
20	D.Kvyat	ToroRosso	↑ 0.431
17	J.Palmer	RenaultRS	↑ 0.499

Quan al pilot que va rebaixar més el seu millor temps de les sessions de pràctiques respecte a la volta de classificació va ser V.Bottas seguit de L.Hamilton. Curiosament, els dos pilots formen part de l'equip MercedesAMG. Tots els pilots, a excepció de D.Kvyat i J.Palmer, van millorar la seva volta de classificació respecte a la simulació.

Com ja s'observa a la part descriptiva, a mesura que passaven les sessions de pràctiques, els temps milloraven. Per tant, es confirma que en la classificació també evolucionen positivament en la majoria dels casos.

A continuació, s'utilitza un altre mètode gràfic per veure si hi ha correlació entre els millors temps analitzats en les sessions de pràctiques (FPBestLap) sobre els temps de classificació (FinalLap).

FIGURA 18: CORRELACIÓ ENTRE FPBESTLAP I FINALLAP



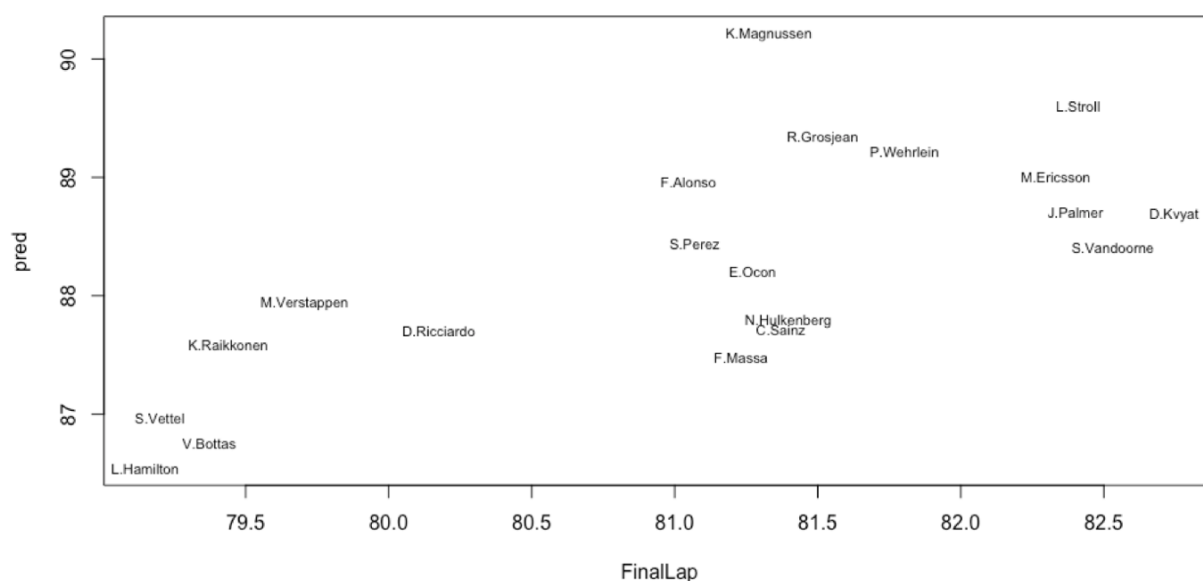
La situació ideal seria tenir tots els pilots formant una diagonal a la Figura 18: Correlació entre FPBestLap i FinalLap. Segons sembla, la correlació entre la volta més ràpida de les sessions i la de classificació és positiva. Per tant, de la informació extreta de les pràctiques es pot arribar a predir l'ordre de la graella, tot i que es difícil preveure el temps exacte de la volta de classificació.

Tal com es va comentar en el moment de la identificació, a primera vista sembla que hi ha dos grups bastant diferenciats. Aquests grups estan formats, per una banda, pels sis primers pilots que condueixen pels equips ScuderiaFerrari, MercedesAMG i RedBull amb dos pilots cadascun. D'altra banda, hi ha la resta de pilots.

El principal problema és la poca diferència que es troba dins de cada subgrup. Per exemple, en el grup dels pilots amb un temps a la volta de classificació entre els 81 i 81,5 segons, n'hi ha vuit. Això provoca que un petit error de pilotatge en el moment de la classificació faci que, en lloc de sortir setè (ja que hi ha molta diferència de rendiment respecte als sis pilots punters) sortís en la catorzena posició, una penalització molt gran per només 0,5 segons de diferència en una volta. Aquesta diferència tan petita és molt difícil de mesurar amb la poca quantitat de dades que es disposa i per tant, complica l'estimació de l'ordre de la graella.

Si ara es fes el mateix procediment però comparant el ritme de cursa estimat (pred) amb la volta de classificació (FinalLap), s'obtindria la Figura 19: Correlació entre pred i FinalLap.

FIGURA 19: CORRELACIÓ ENTRE PRED I FINALLAP



Tal com s'ha observat prèviament a la Figura 18: Correlació entre FPBestLap i FinalLap, la correlació sembla positiva en els sis monoplaques punters. En canvi, en la Figura 19: Correlació entre pred i FinalLap, per la part del darrere de la graella, aquesta correlació ja no és tan clara. El que això indica és que la velocitat que presenta un monoplaça a una volta no ha de ser la mateixa que la que en el moment de la cursa, fent que vagi perdent posicions a mesura que passen les voltes. Aquest seria el cas, per exemple, de F. Alonso que va classificar molt bé, però el seu ritme de cursa estimat no sembla suficient per mantenir aquesta setena posició inicial. El cas contrari podria ser el del pilot F. Massa, que no va classificar gaire bé, però el seu ritme de cursa estimat sembla molt bo. De manera global, el ritme de cursa d'aquest segon grup és bastant semblant, a excepció del pilot K. Magnussen que segurament acabi dels últims. Tot fa indicar que la resta de posicions es decidiran per petits detalls.

Abans de passar a comparar els resultats, s'ha de dir que en una cursa de Formula 1 succeeixen imprevistos difícils de preveure, com per exemple: accidents, errors de pilotatge, diverses estratègies, punxades de pneumàtics o inclòs el simple fet que pugui sortir el *safety car* pot fer variar les estratègies dels equips radicalment.

L'estimació de la posició final per a una cursa sense accidents, només tenint en compte el ritme mostrat pels pilots en les sessions de pràctiques i en una situació en la qual avançar un cotxe més lent no suposés cap pèrdua de temps, s'obtingria la Taula 8: Estimació de l'ordre final de la cursa (sense incidents).

TAULA 8: ESTIMACIÓ DE L'ORDRE FINAL DE LA CURSA (SENSE INCIDENTS)

Posició	Pilot	Posició	Pilot
1	L.Hamilton	11	S.Vandoorne
2	V.Bottas	12	S.Perez
3	S.Vettel	13	D.Kvyat
4	F.Massa	14	J.Palmer
5	K.Raikkonen	15	F.Alonso
6	D.Ricciardo	16	M.Ericsson
7	C.Sainz	17	P.Wehrlein
8	N.Hulkenberg	18	R.Grosjean
9	M.Verstappen	19	L.Stroll
10	E.Ocon	20	K.Magnussen

Si s'assenyalen en aquest resultat final estimat els incidents que va haver-hi durant el Gran Premi de Catalunya, s'obté la Taula 9: Estimació del resultat final de la cursa indicant incidents.

TAULA 9: ESTIMACIÓ DEL RESULTAT FINAL DE LA CURSA INDICANT INCIDENTS

Posició	Pilot	Posició	Pilot
1	L.Hamilton	11	S.Vandoorne
2	V.Bottas	12	S.Perez
3	S.Vettel	13	D.Kvyat
4	F.Massa	14	J.Palmer
5	K.Raikkonen	15	F.Alonso
6	D.Ricciardo	16	M.Ericsson
7	C.Sainz	17	P.Wehrlein
8	N.Hulkenberg	18	R.Grosjean
9	M.Verstappen	19	L.Stroll
10	E.Ocon	20	K.Magnussen

Els pilots marcats en vermell no van acabar la cursa i els que estan marcats en taronja van patir algun tipus d'incident, com una punxada al pneumàtic o una sortida de pista. Per tant, si es prescindís d'aquests set pilots que no van poder córrer amb normalitat a la cursa, s'obté la Taula 10: Estimació del resultat final de la cursa corregit per alguns incidents.

TAULA 10: ESTIMACIÓ DEL RESULTAT FINAL DE LA CURSA CORREGIT PER ALGUNS INCIDENTS

Posició	Pilot	Posició	Pilot
1	L.Hamilton	8	D.Kvyat
2	S.Vettel	9	J.Palmer
3	D.Ricciardo	10	M.Ericsson
4	C.Sainz	11	P.Wehrlein
5	N.Hulkenberg	12	R.Grosjean
6	E.Ocon	13	L.Stroll
7	S.Perez		

Resultat final real de la cursa del Gran Premi de Catalunya 2017:

TAULA 11: RESULTAT FINAL REAL DE LA CURSA

Posició	Pilot	Posició	Pilot
1	L.Hamilton	9	D.Kvyat
2	S.Vettel	10	R.Grosjean
3	D.Ricciardo	11	M.Ericsson
4	S.Perez	12	F.Alonso
5	E.Ocon	13	F.Massa
6	N.Hulkenberg	14	K.Magnussen
7	C.Sainz	15	J.Palmer
8	P.Wehrlein	16	L.Stroll

En el moment que es corregeix el resultat final per alguns dels incidents que van succeir a la cursa, tot i que encara és difícil magnificar l'efecte del *virtual safety car* que va aparèixer en alguns moments de la cursa, s'aconsegueix una bona aproximació i, per tant, el model prediu bé el resultat final a partir de les simulacions de cursa obtingudes de les sessions de pràctiques.

En resum, s'aconsegueix predir l'ordre del pòdium de la cursa, els set primers pilots i, per a la resta, l'error més gran de l'estimació és de tres posicions respecte al real. Aquests pilots són P.Wehrlein i J.Palmer. Wehrlein va ser l'únic en optar per una estratègia d'una parada, que era l'òptima tal com s'ha comentat prèviament. A més, va tenir la sort de parar en període de *virtual safety car* en la volta trenta-tres quan portava el compost més tou. El número de voltes òptim pel compost tou eren trenta-cinc. Això li va permetre perdre menys temps i guanyar algunes posicions. Per contra, J.Palmer va fer l'oposat: va parar tres vegades

i, això, li va suposar perdre un temps que després no va ser capaç de recuperar. La resta de desviacions de posicions són mínimes, tot i les incidències aparegudes durant la cursa.

Per tant, tal com s'ha predit amb les dades de les sessions de pràctiques, després de la identificació de les simulacions de cursa i interpretant els resultats del model obtingut, es confirma, gràcies a P.Wehrlein, que l'estratègia d'una parada era l'òptima per completar en el mínim de temps les seixanta-sis voltes que completen el Gran Premi de Catalunya.

5. CONCLUSIONS

Al començament del treball es van plantejar una sèrie d'objectius d'estudi. L'objectiu principal era fer una prognosi del resultat final de la cursa a partir de la informació extreta dels temps de les tres sessions de pràctiques. Per tant, amb aquestes dades es pretenia identificar les simulacions de cursa realitzades per cada pilot i crear un model estadístic que fos capaç de predir el ritme de cursa de cada monoplaça i anticipar el resultat final, a més de calcular la millor estratègia de parades en boxes a partir de la modelització del nivell de degradació dels diferents compostos disponibles. Entre els objectius secundaris es trobava també la identificació de la simulació de classificació i, finalment, la categorització dels pilots i equips de la graella.

Inicialment, es va realitzar un procés de construcció de la base de dades, ja que aquesta només estava disponible en un format incompatible amb l'RStudio. D'altra banda, hi havia certes manques d'informació, sobretot les referides als compostos emprats en cada volta. Un cop es va tenir disponible la base de dades en format Excel, es va prosseguir amb l'anàlisi exploratori de les dades.

Arran de l'anàlisi descriptiva de les dades, s'han extret algunes conclusions que s'han tornat més sòlides després de dur a terme el model mixte. L'exemple més clar s'ha pogut observar en les voltes més ràpides de cada sessió. A mesura que s'avançava en els entrenaments lliures, els monoplaces anaven cada vegada més ràpid, fet que feia pensar que en el moment de la classificació real poguessin rebaixar encara més els millors temps marcats en els entrenaments. Finalment, aquesta hipòtesi es va confirmar. També es va observar que la simulació de classificació es realitzava a l'última sessió de pràctiques (FP3), i hi participaven tots els pilots amb el compost més ràpid dels tres (S). Quant al resultat estimat de la graella de sortida, presenta una coherència acceptable on es poden diferenciar dos grups: els equips punters (MercedesAMG, ScuderiaFerrari i RedBull) i la resta d'equips. Aquests dos grups són heterogenis entre ells, però homogenis dins de cadascun, fet que complicava l'ordenació dels pilots.

Pel que fa a la informació referida a la cursa, les dades presentaven un gran soroll. Aquest soroll podia provenir d'errors de pilotatge o voltes en situació de bandera groga en la qual tots els monoplaces han de minorar la velocitat, entre d'altres. El que no s'esperava era trobar patrons en els quals hi hagués una volta molt lenta seguida d'una molt ràpida. Després d'una exhaustiva investigació, es va concloure que aquests patrons tan anòmals eren provocats pel tipus de motor dels monoplaces, ja que aquest, al ser híbrids, necessitaven una volta per recarregar les bateries. Clarament, aquestes voltes no són

representatives del ritme de cursa. Per tant, l'única forma d'identificar la sèrie de voltes que pretenen reflectir el ritme de cursa, és creant uns criteris que permetin aïllar les dades amb les quals es vol treballar. És a dir, netejant tot el soroll provocat per aquestes observacions.

Un cop la base de dades es troba neta de distorsions, ha calgut utilitzar una metodologia que modelitzi dades longitudinals. A causa del fet que hi ha poques dades per cada pilot, no es pot afirmar que aquestes pertanyin a sèries temporals. Per tant, s'apliquen models mixtes, els més apropiats donat aquest context.

Els models mixtes han estat efectius a l'hora de modelitzar i quantificar l'efecte de la degradació dels pneumàtics en el temps per volta. Pel que fa a les dades del Gran Premi del 2017, aquestes han permès identificar un model amb degradació comuna pels diferents compostos, però amb un millor rendiment pels tous (S). També s'ha aplicat aquesta metodologia a les dades del Gran Premi del 2016, obtenint, en aquest cas, un resultat que permet caracteritzar millor la degradació i el rendiment dels pneumàtics.

Es pot observar que en els dos casos (anys 2016/17) hi ha un salt de rendiment significatiu entre compostos. Tot i això, la degradació presentada en l'any 2016 ja no es comporta de la mateixa manera pels diferents compostos. En el cas del 2016, el compost més ràpid es degrada més i el compost més dur no presenta cap desgast. A partir de la setena o vuitena voltes hi ha un punt d'inflexió en el qual el dur passa a ser el més competitiu dels tres.

Una de les utilitats que s'han pogut derivar de la metodologia és la possibilitat de determinar una estratègia de parades òptima per realitzar el canvi obligatori de pneumàtics. Només ha calgut definir una funció objectiu que tingués en compte el nivell de degradació i el seu efecte sobre el temps de cada volta. Optimitzant aquesta funció s'ha trobat que pel Gran Premi del 2017 s'haurien de fer 31 voltes amb el mitjà (M) i 35 amb el tou (S).

Respecte als models resultants, aquests s'han validat correctament amb l'anàlisi de residus i tractament de atípics. A partir d'aquests models, s'han fet prediccions dels temps de cursa per a diferents voltes (2, 10, 64) i per a les posicions finals.

A l'hora de comparar els resultats amb els valors reals posteriors, s'ha vist una certa coherència global. Tot i això, s'ha de tenir en compte que en la cursa va haver-hi diferents incidents que poden desvirtuar la comparativa, com per exemple: accidents, punxades i la sortida de *safety cars*. Aquests provoquen variacions en les estratègies. Finalment, els resultats obtinguts s'han contrastat amb un expert i, aquests, s'han trobat molt raonables tenint en compte totes les incidències produïdes durant la cursa. Resumint, s'ha aconseguit predir l'ordre del pòdium de la cursa, els set primers pilots i, de la resta de pilots, que l'error més gran de l'estimació és de tres posicions respecte al real. Aquests pilots són P.Wehrlein i J.Palmer. Wehrlein va ser l'únic en optar per una estratègia d'una parada, que era l'òptima,

tal com s'ha comentat prèviament. Gràcies a aquest moviment, el resultat real està tres posicions per sota de l'estimat, d'onze a vuitè. Per contra, J.Palmer va fer l'oposat: va parar tres vegades, el que li va suposar perdre un temps que després no va ser capaç de recuperar. També s'ha determinat que la FP2 és la sessió de lliures en la que els pilots realitzen la simulació de cursa.

Els efectes aleatoris del model mixte permeten categoritzar per cada un dels pilots i equips. Només es van trobar diferències significatives entre els dos pilots de l'equip Williams i entre els equips MercedesAMG amb Haas i Sauber i ScuderiaFerrari amb Haas. També s'han avaluat altres efectes com el tipus de motor de cada monoplaça sense que apareguin diferències significatives però, aquestes, poden estar afectades per altres factors confusos molt difícils o gairebé impossibles de mesurar, com l'efecte de les suspensions o l'aerodinàmica.

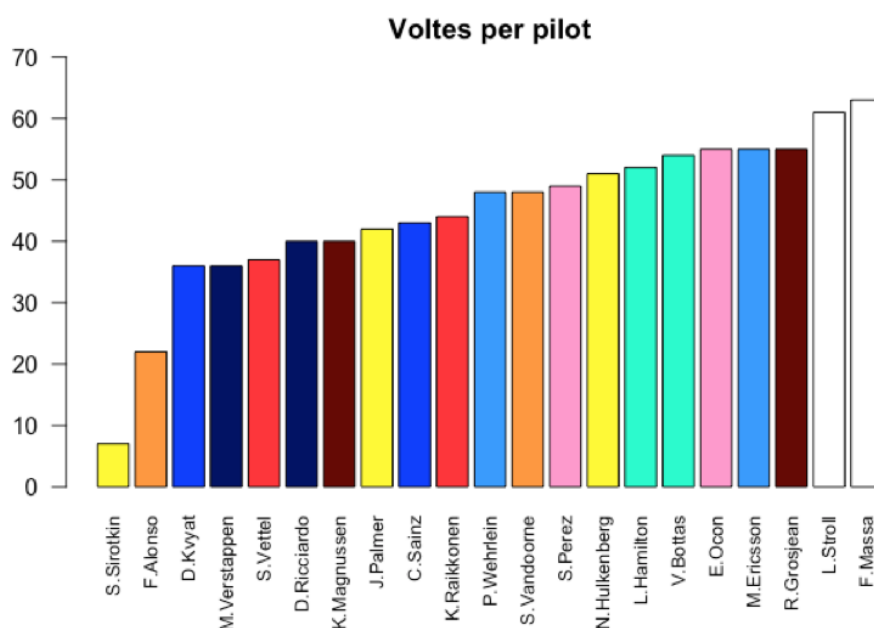
En definitiva, fer aquest treball m'ha permès aplicar eines estadístiques en unes dades complicades de tractar per el gran soroll que presenten. Una vegada vaig aconseguir un codi el suficientment eficaç a l'hora d'identificar les sèries de voltes pertanyents a una simulació de cursa, la hipòtesis de la utilització de sèries temporals va quedar descartada per la poca quantitat de dades netes. Finalment, em vaig decantar per la utilització de models mixtes. Aquest tipus de models no es van treballar en profunditat en l'assignatura "Models generalitzats", per tant, els vaig utilitzar de manera autònoma passant per una fase d'autoaprenentatge. A més, vaig utilitzar noves eines gràfiques molt sofisticades i potents com *plotly* i *formattable*. En aquest informe també he utilitzat tècniques estadístiques de l'assignatura "Disseny d'experiments".

6. ANNEX

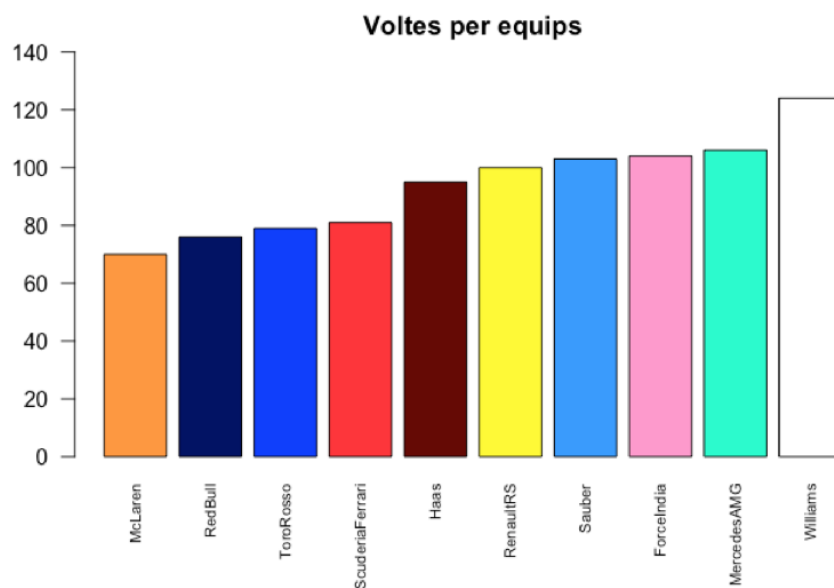
Annex A: GP 2017

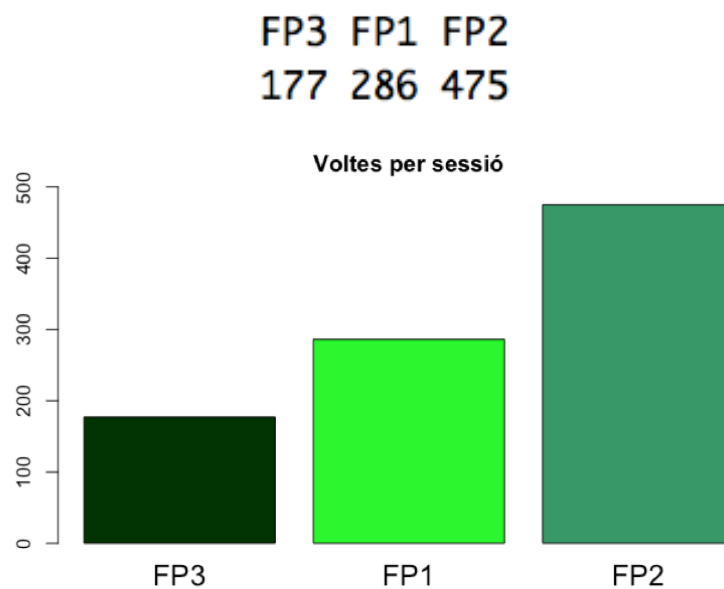
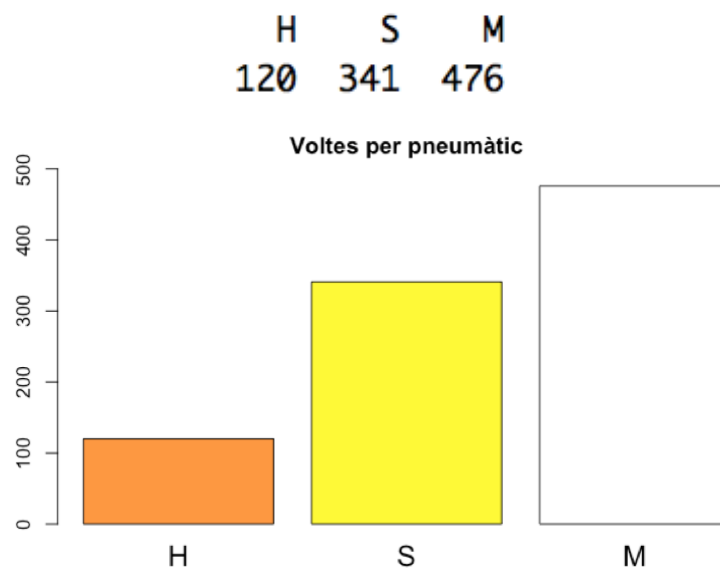
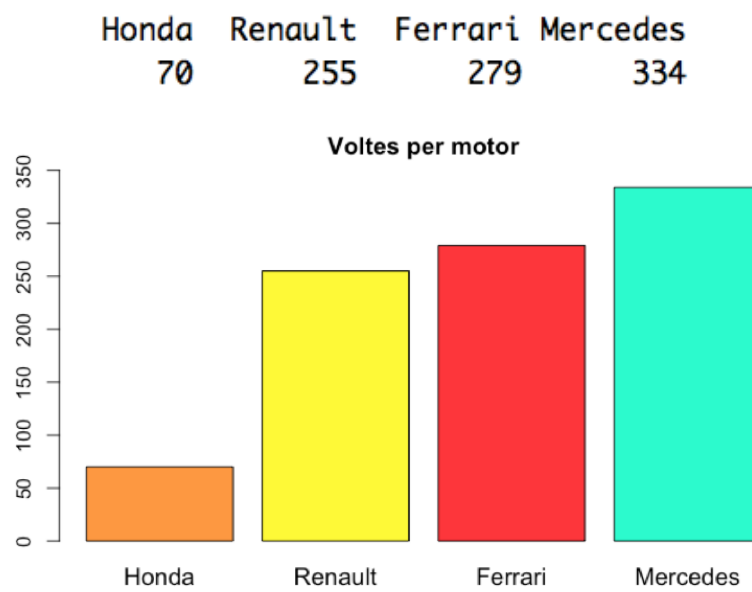
Descriptiva Univariada

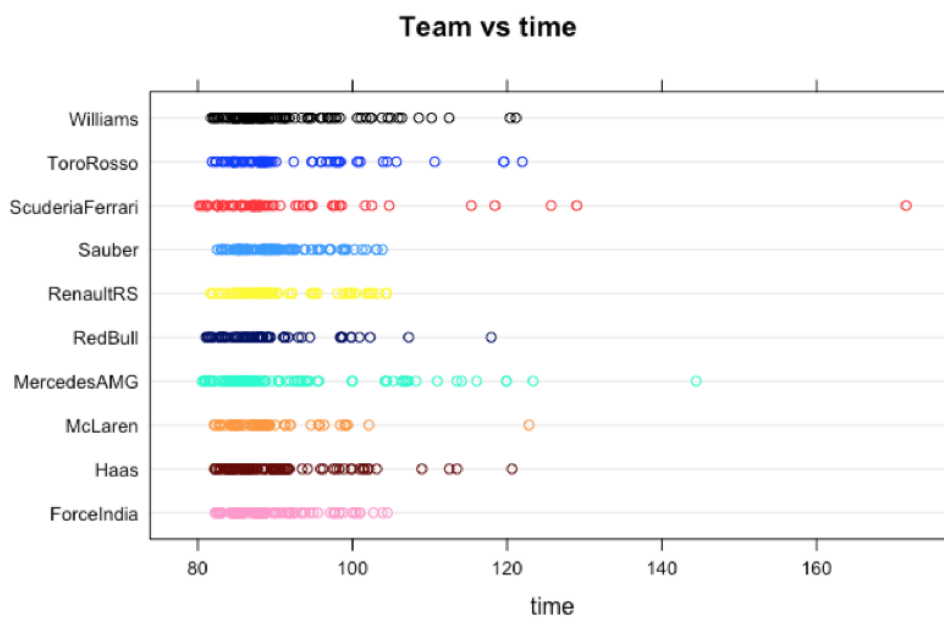
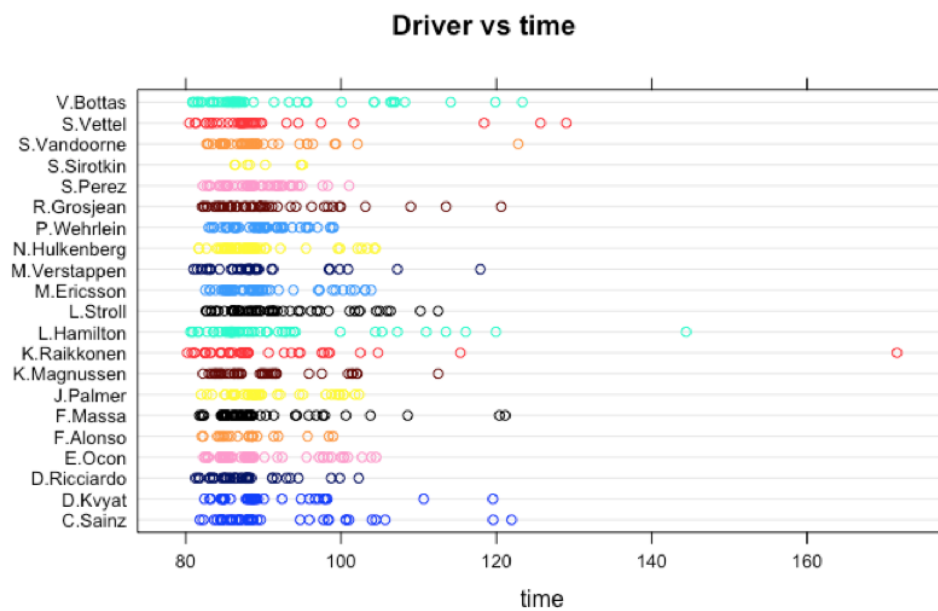
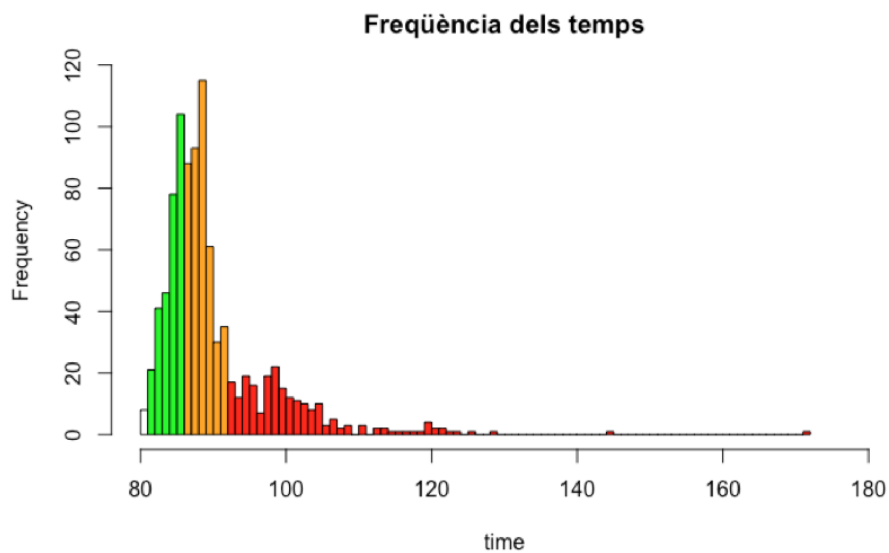
S.Sirotkin	F.Alonso	D.Kvyat	M.Verstappen	S.Vettel	D.Ricciardo
7	22	36	36	37	40
K.Magnussen	J.Palmer	C.Sainz	K.Raikkonen	P.Wehrlein	S.Vandoorne
40	42	43	44	48	48
S.Perez	N.Hulkenberg	L.Hamilton	V.Bottas	E.Ocon	M.Ericsson
49	51	52	54	55	55
R.Grosjean	L.Stroll	F.Massa			
55	61	63			



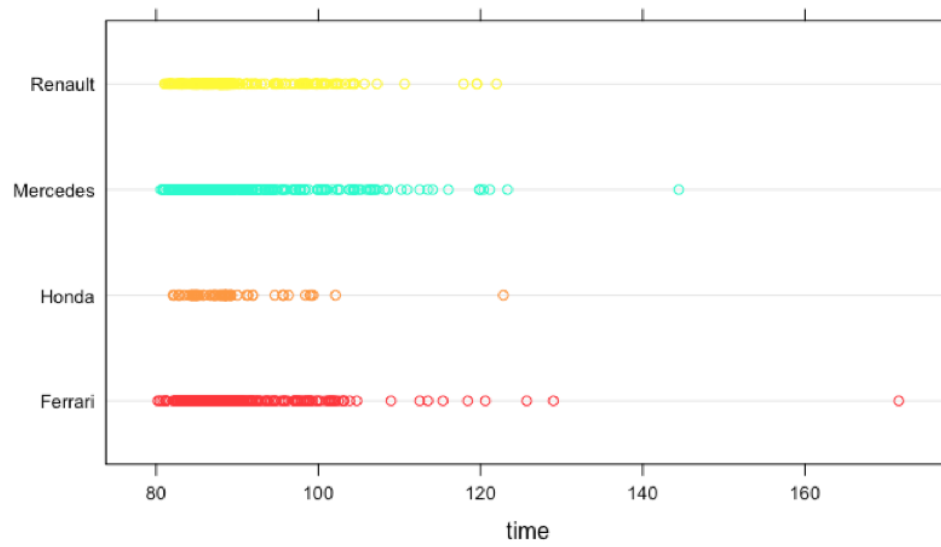
McLaren	Red Bull	Toro Rosso	Scuderia Ferrari	Haas
70	76	79	81	95
Renault RS	Sauber	Force India	Mercedes AMG	Williams
100	103	104	106	124



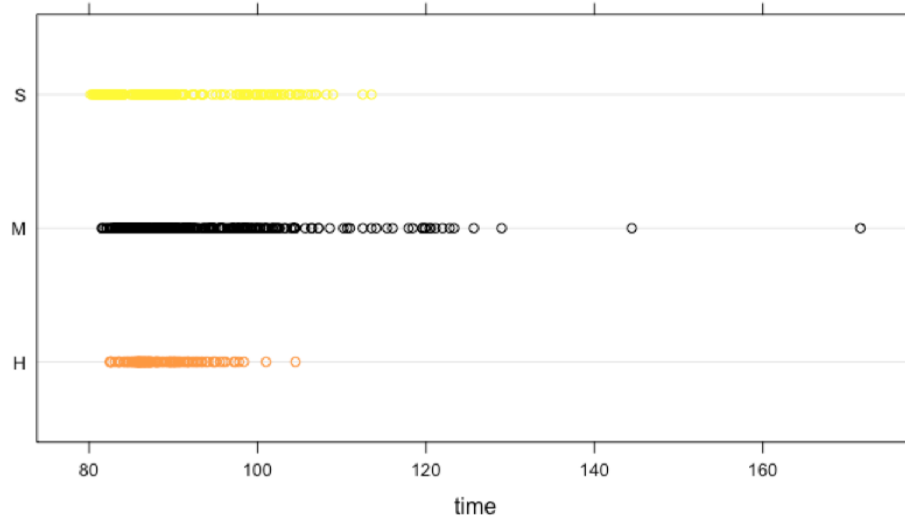




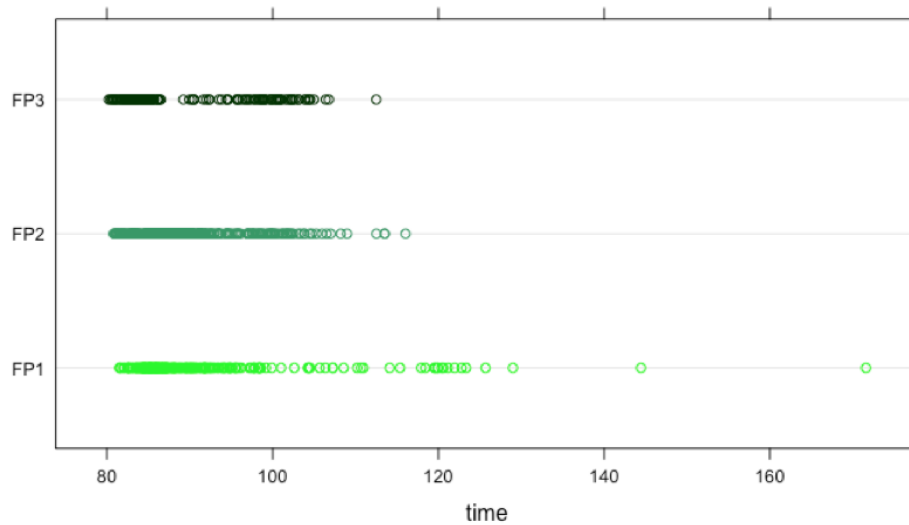
Engine vs time



Tyre vs time

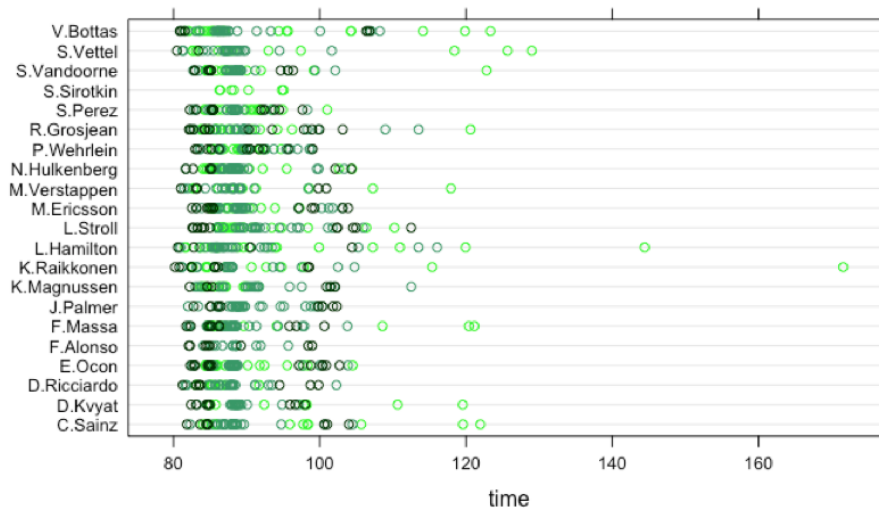


Session vs time

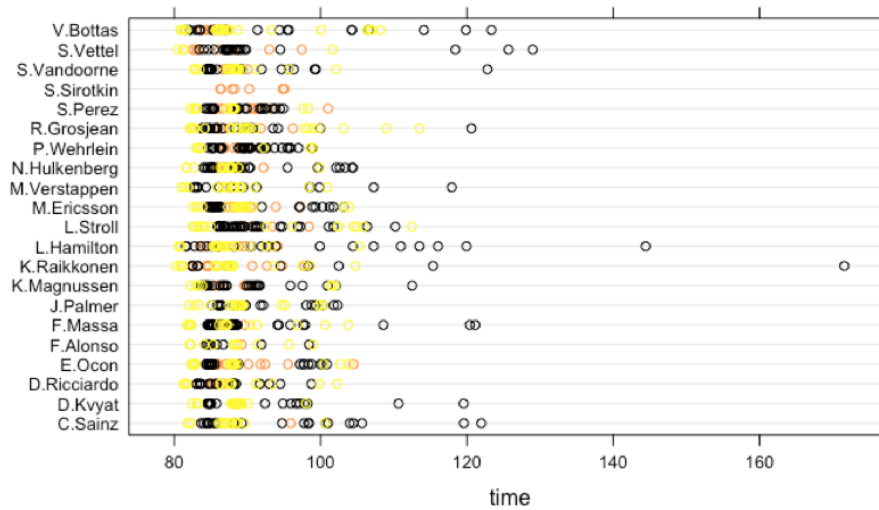


Descriptiva bivariada

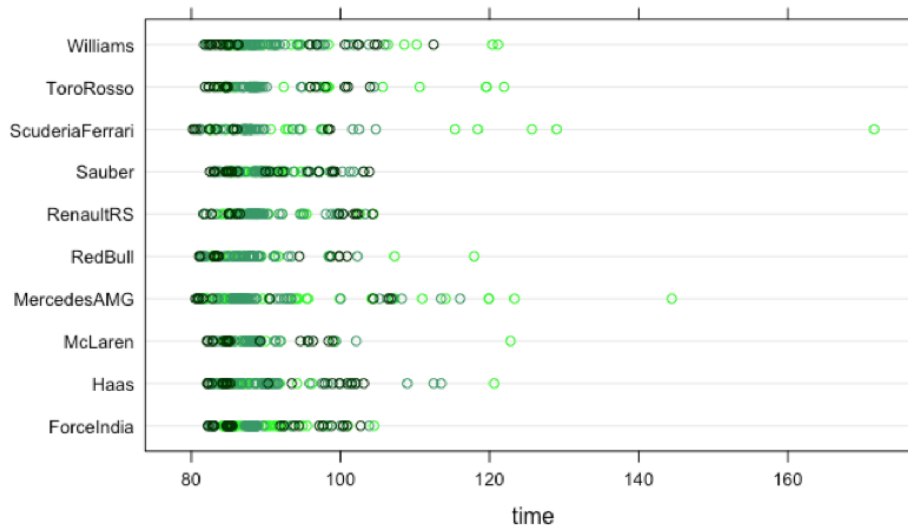
Driver vs time segons Session



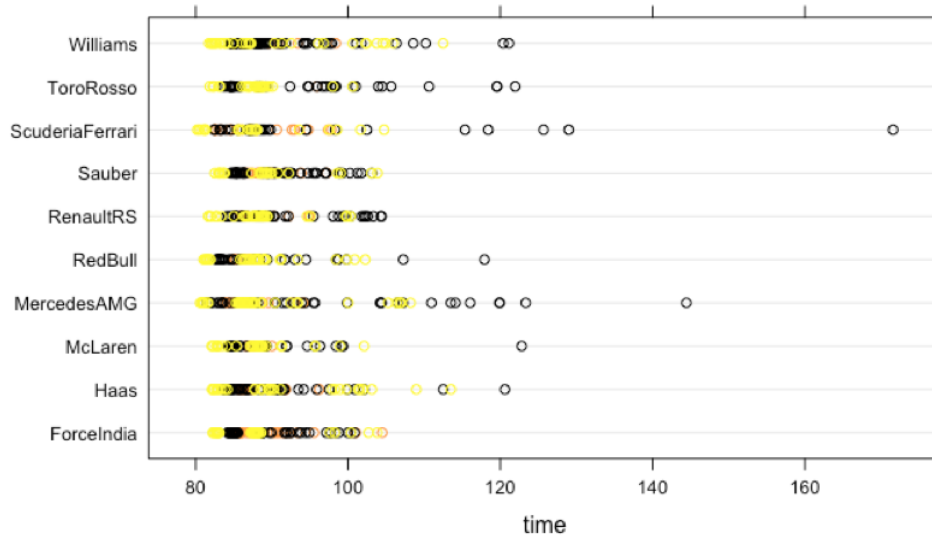
Driver vs time segons Tyre



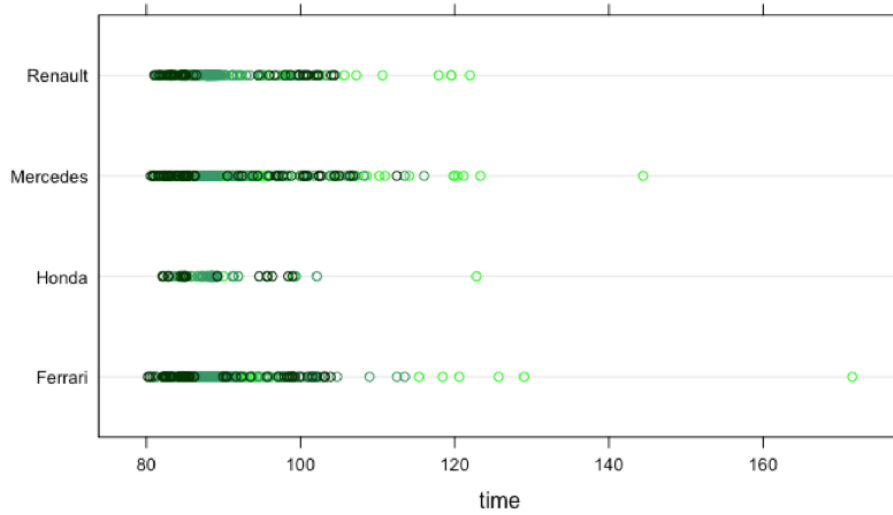
Team vs time segons Session



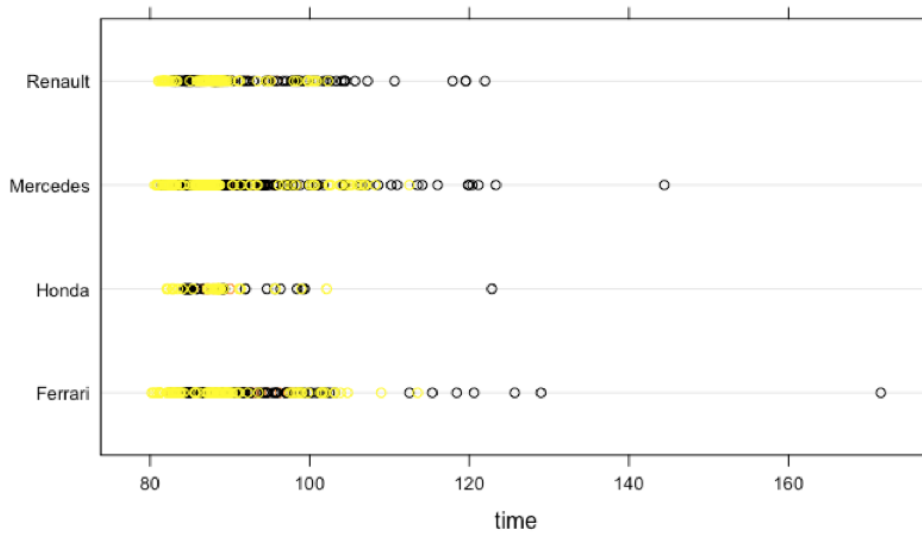
Team vs time segons Tyre



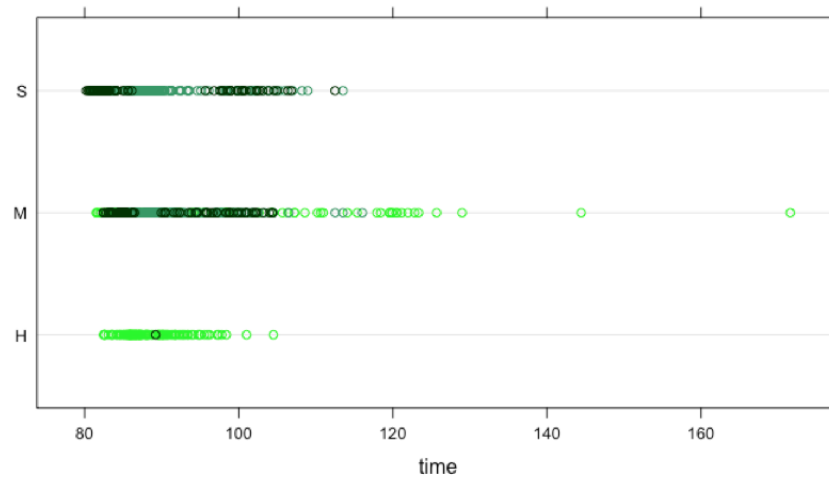
Engine vs time segons Session



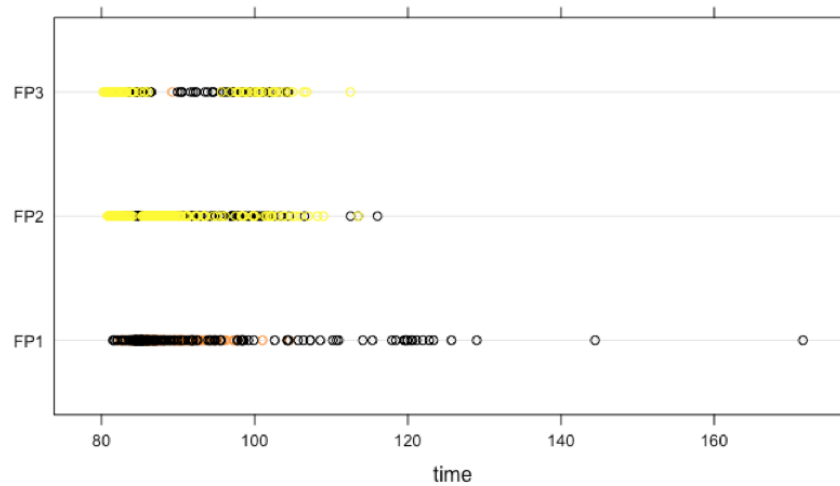
Engine vs time segons Tyre



Tyre vs time segons Session

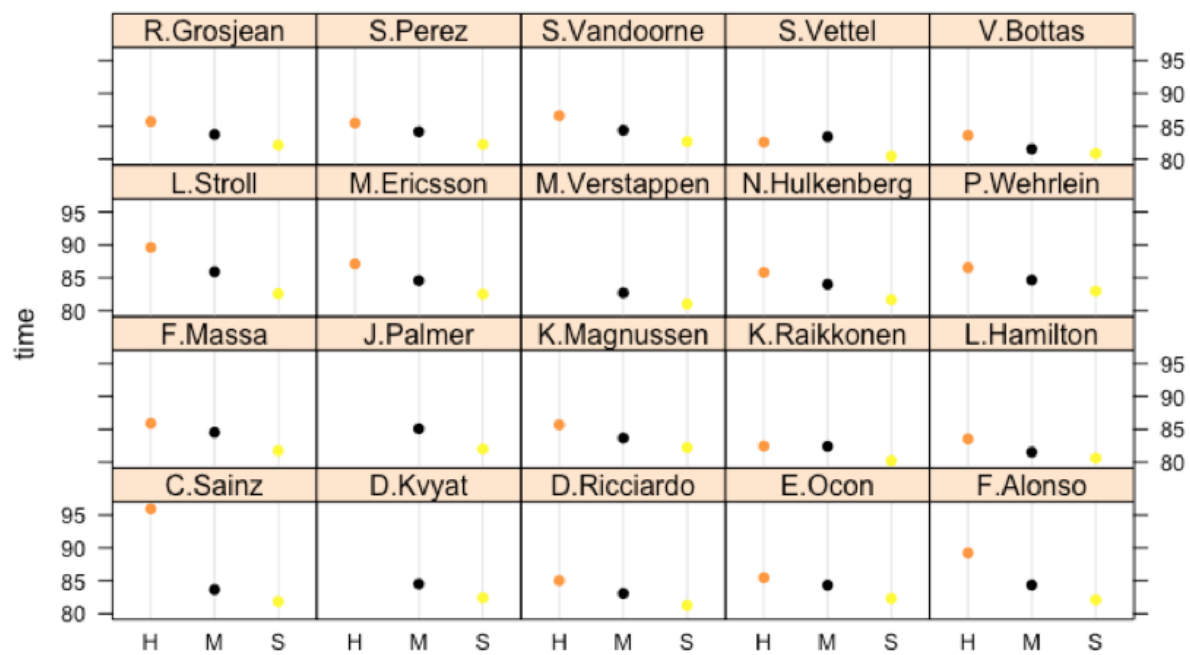


Session vs time segons Tyre

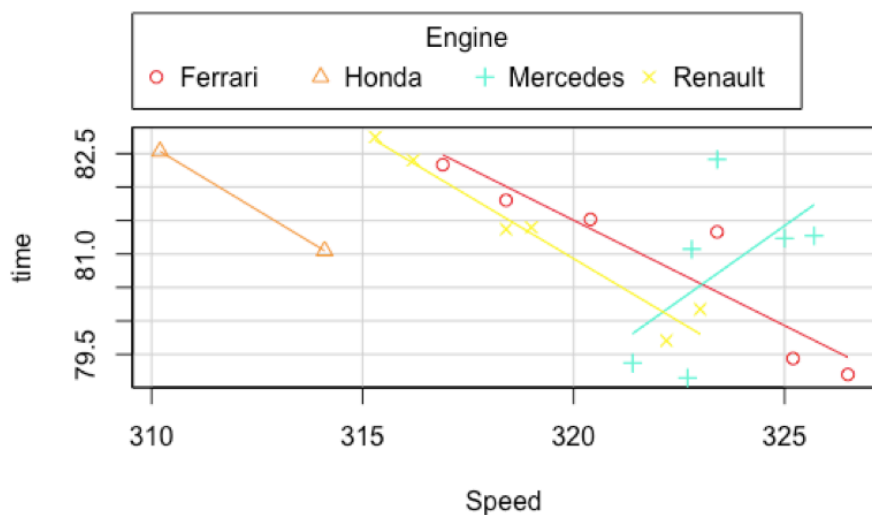


Resultats

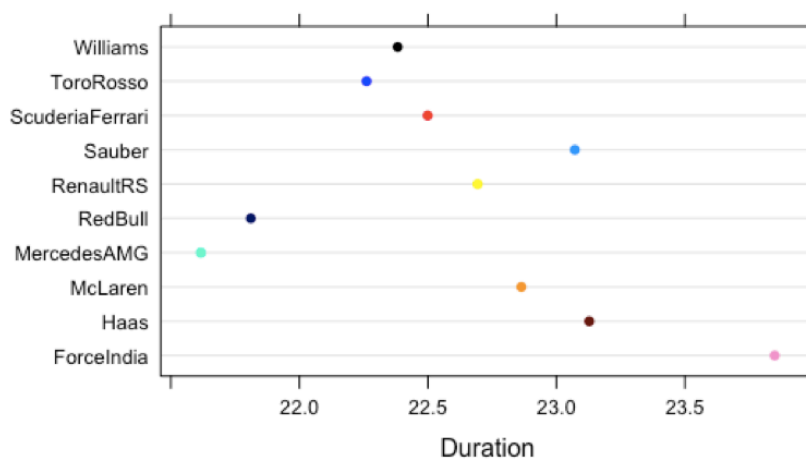
	GridPosition	Number	Driver	Team	Engine Tyre	Time Session	minuts	segons	time
784	1	7	K.Raikkonen	ScuderiaFerrari	Ferrari	S 1:20.214	FP3	1 20.214	80.214
780	2	5	S.Vettel	ScuderiaFerrari	Ferrari	S 1:20.456	FP3	1 20.456	80.456
914	3	44	L.Hamilton	MercedesAMG	Mercedes	S 1:20.595	FP3	1 20.595	80.595
923	4	77	V.Bottas	MercedesAMG	Mercedes	S 1:20.868	FP3	1 20.868	80.868
909	5	33	M.Verstappen	RedBull	Renault	S 1:21.025	FP3	1 21.025	81.025
777	6	3	D.Ricciardo	RedBull	Renault	S 1:21.249	FP3	1 21.249	81.249
879	7	27	N.Hulkenberg	RenaultRS	Renault	S 1:21.67	FP3	1 21.670	81.670
855	8	19	F.Massa	Williams	Mercedes	S 1:21.746	FP3	1 21.746	81.746
921	9	55	C.Sainz	ToroRosso	Renault	S 1:21.835	FP3	1 21.835	81.835
593	10	30	J.Palmer	RenaultRS	Renault	S 1:21.992	FP2	1 21.992	81.992
833	11	14	F.Alonso	McLaren	Honda	S 1:22.093	FP3	1 22.093	82.093
797	12	8	R.Grosjean	Haas	Ferrari	S 1:22.128	FP3	1 22.128	82.128
864	13	20	K.Magnussen	Haas	Ferrari	S 1:22.214	FP3	1 22.214	82.214
825	14	11	S.Perez	ForceIndia	Mercedes	S 1:22.237	FP3	1 22.237	82.237
900	15	31	E.Ocon	ForceIndia	Mercedes	S 1:22.297	FP3	1 22.297	82.297
871	16	26	D.Kvyat	ToroRosso	Renault	S 1:22.391	FP3	1 22.391	82.391
813	17	9	M.Ericsson	Sauber	Ferrari	S 1:22.513	FP3	1 22.513	82.513
843	18	18	L.Stroll	Williams	Mercedes	S 1:22.574	FP3	1 22.574	82.574
291	19	2	S.Vandoorne	McLaren	Honda	S 1:22.693	FP2	1 22.693	82.693
936	20	94	P.Wehrlein	Sauber	Ferrari	S 1:22.974	FP3	1 22.974	82.974
237	21	46	S.Sirotkin	RenaultRS	Renault	H 1:26.293	FP1	1 26.293	86.293



GridPosition	BesttimeSoft	BesttimeMedium	BesttimeHard
1	K.Raikkonen	L.Hamilton	K.Raikkonen
2	S.Vettel	V.Bottas	S.Vettel
3	L.Hamilton	K.Raikkonen	L.Hamilton
4	V.Bottas	M.Verstappen	V.Bottas
5	M.Verstappen	D.Ricciardo	D.Ricciardo
6	D.Ricciardo	S.Vettel	E.Ocon
7	N.Hulkenberg	C.Sainz	S.Perez
8	F.Massa	K.Magnussen	K.Magnussen
9	C.Sainz	R.Grosjean	R.Grosjean
10	J.Palmer	N.Hulkenberg	N.Hulkenberg
11	F.Alonso	S.Perez	F.Massa
12	R.Grosjean	E.Ocon	P.Wehrlein
13	K.Magnussen	F.Alonso	S.Vandoorne
14	S.Perez	S.Vandoorne	M.Ericsson
15	E.Ocon	D.Kvyat	F.Alonso
16	D.Kvyat	F.Massa	L.Stroll
17	M.Ericsson	M.Ericsson	C.Sainz
18	L.Stroll	P.Wehrlein	<NA>
19	S.Vandoorne	J.Palmer	<NA>
20	P.Wehrlein	L.Stroll	<NA>



Mitjana dels Pit Stops segons Team



```
> summary(modT<-lm(time~Team,FastLaps))
```

Call:

```
lm(formula = time ~ Team, data = FastLaps)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.6483	-0.2305	-0.0300	0.1365	2.9747

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	82.2670	0.8036	102.378	<2e-16 ***
TeamHaas	-0.0960	1.1364	-0.084	0.934
TeamMcLaren	0.1260	1.1364	0.111	0.914
TeamMercedesAMG	-1.5355	1.1364	-1.351	0.204
TeamRedBull	-1.1300	1.1364	-0.994	0.341
TeamRenaultRS	1.0513	1.0374	1.013	0.333
TeamSauber	0.4765	1.1364	0.419	0.683
TeamScuderiaFerrari	-1.9320	1.1364	-1.700	0.117
TeamToroRosso	-0.1540	1.1364	-0.136	0.895
TeamWilliams	-0.1070	1.1364	-0.094	0.927

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.136 on 11 degrees of freedom

Multiple R-squared: 0.5471, Adjusted R-squared: 0.1765

F-statistic: 1.476 on 9 and 11 DF, p-value: 0.2673

```
> summary(modE<-lm(time~Engine,FMHR))
```

Call:
lm(formula = time ~ Engine, data = FMHR)

Residuals:

783	779	913	922	878	592	832	290
-0.1210	0.1210	-0.1365	0.1365	-0.1610	0.1610	-0.3000	0.3000

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	80.3350	0.1931	415.967	2e-10 ***
EngineHonda	2.0580	0.2731	7.535	0.00166 **
EngineMercedes	0.3965	0.2731	1.452	0.22022
EngineRenault	1.4960	0.2731	5.477	0.00541 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2731 on 4 degrees of freedom
Multiple R-squared: 0.9482, Adjusted R-squared: 0.9093
F-statistic: 24.39 on 3 and 4 DF, p-value: 0.00495

```
> summary(modTy<-lm(time~Tyre,FastLapsSMH))
```

Call:
lm(formula = time ~ Tyre, data = FastLapsSMH)

Residuals:

Min	1Q	Median	3Q	Max
-3.1982	-0.6644	0.1759	0.6989	3.9558

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	85.6542	0.3381	253.369	< 2e-16 ***
TyreM	-1.8401	0.4536	-4.057	0.000164 ***
TyreS	-3.8660	0.4536	-8.524	1.65e-11 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.352 on 53 degrees of freedom
Multiple R-squared: 0.5804, Adjusted R-squared: 0.5646
F-statistic: 36.66 on 2 and 53 DF, p-value: 1.01e-10

```
> summary(modS<-lm(time~Session,fp))
```

Call:
lm(formula = time ~ Session, data = fp)

Residuals:

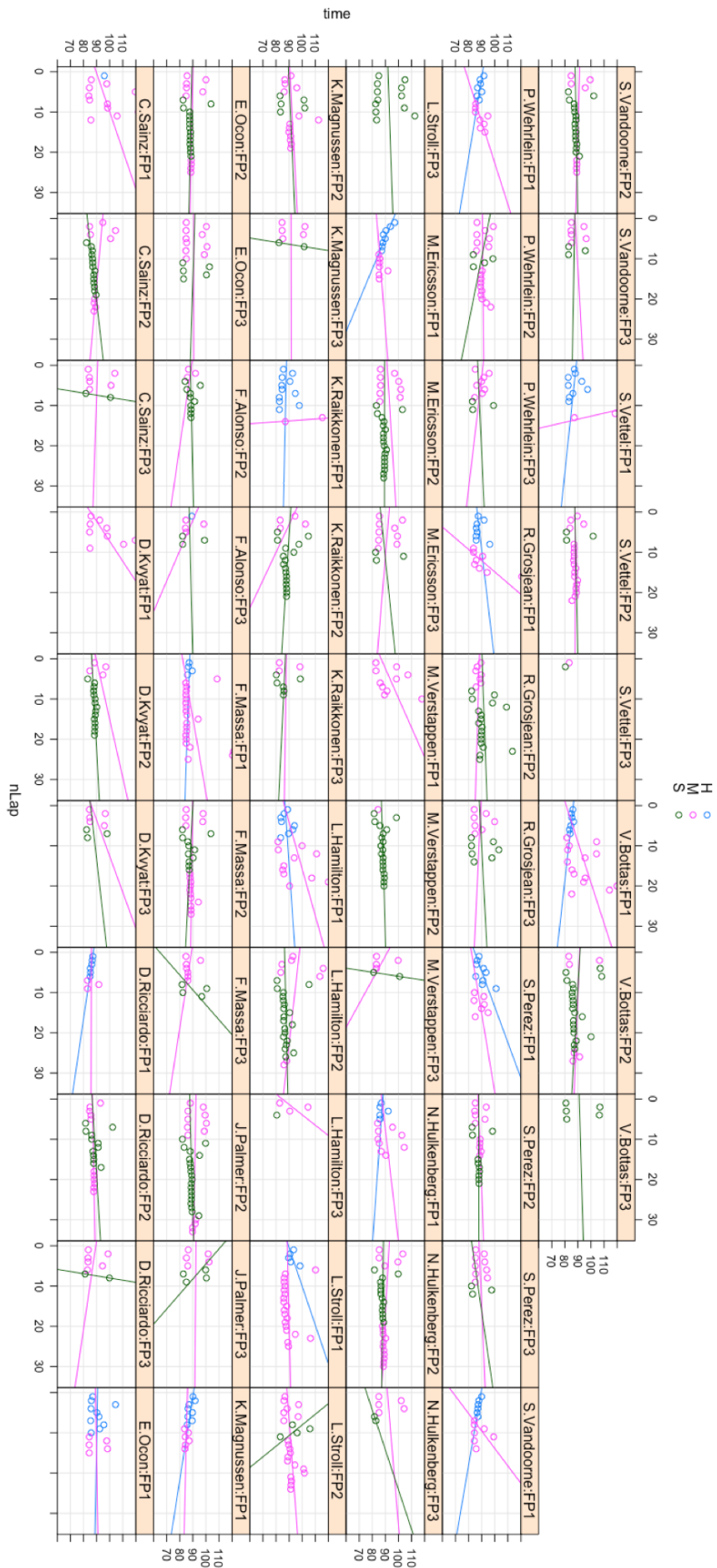
Min	1Q	Median	3Q	Max
-9.717	-4.849	-1.663	1.419	80.343

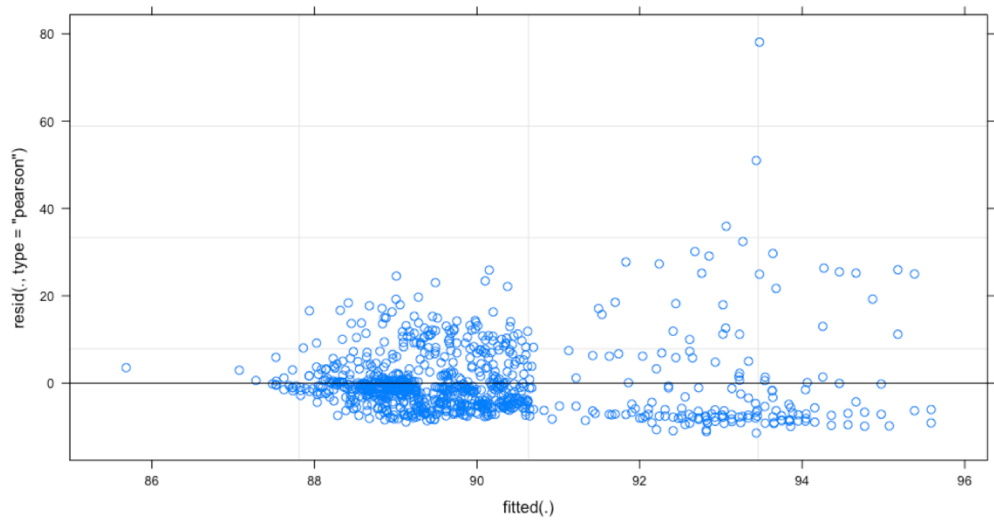
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	91.2372	0.4756	191.827	< 2e-16 ***
SessionFP2	-1.8029	0.5988	-3.011	0.00268 **
SessionFP3	-1.6360	0.7626	-2.145	0.03218 *

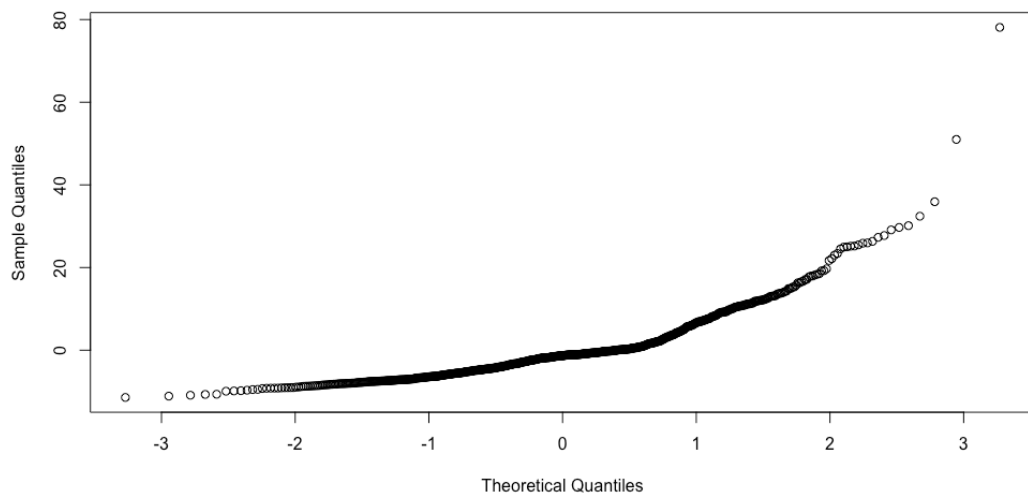
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.93 on 927 degrees of freedom
Multiple R-squared: 0.01028, Adjusted R-squared: 0.008147
F-statistic: 4.815 on 2 and 927 DF, p-value: 0.008309

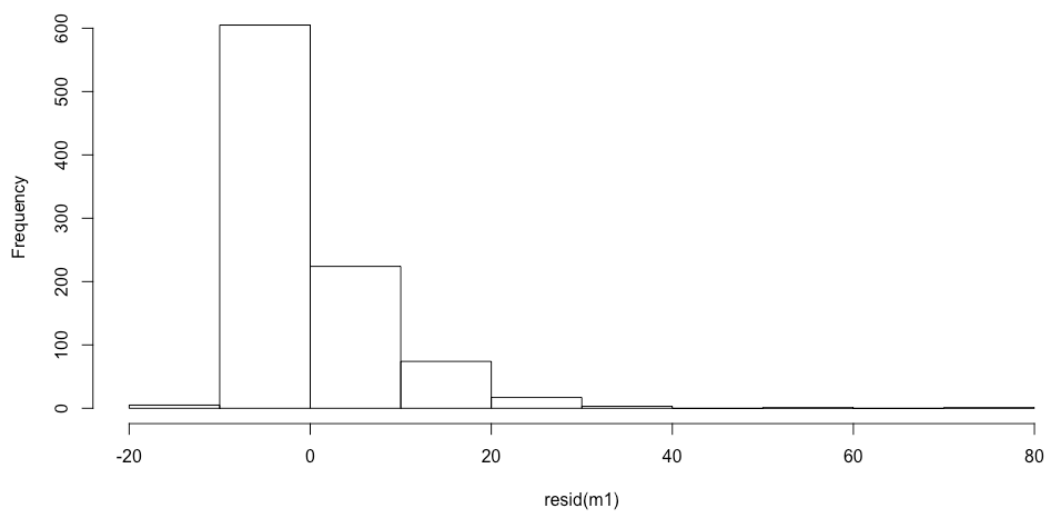


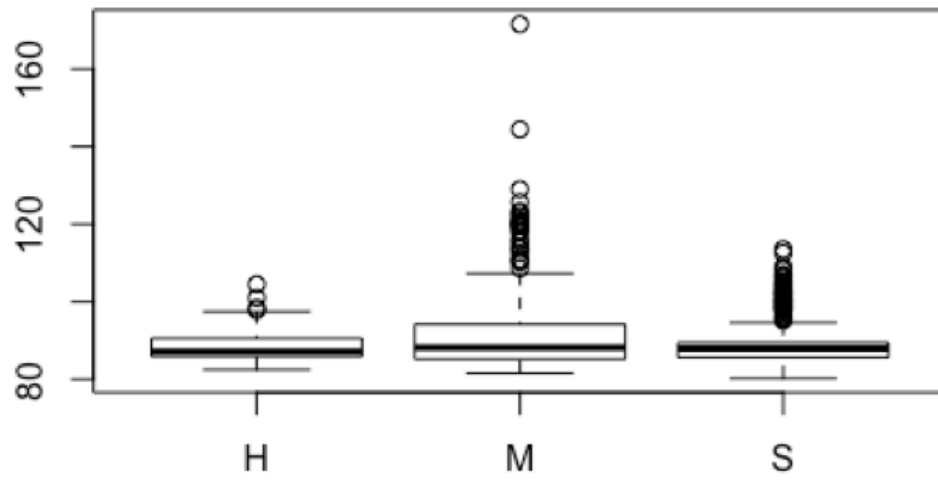


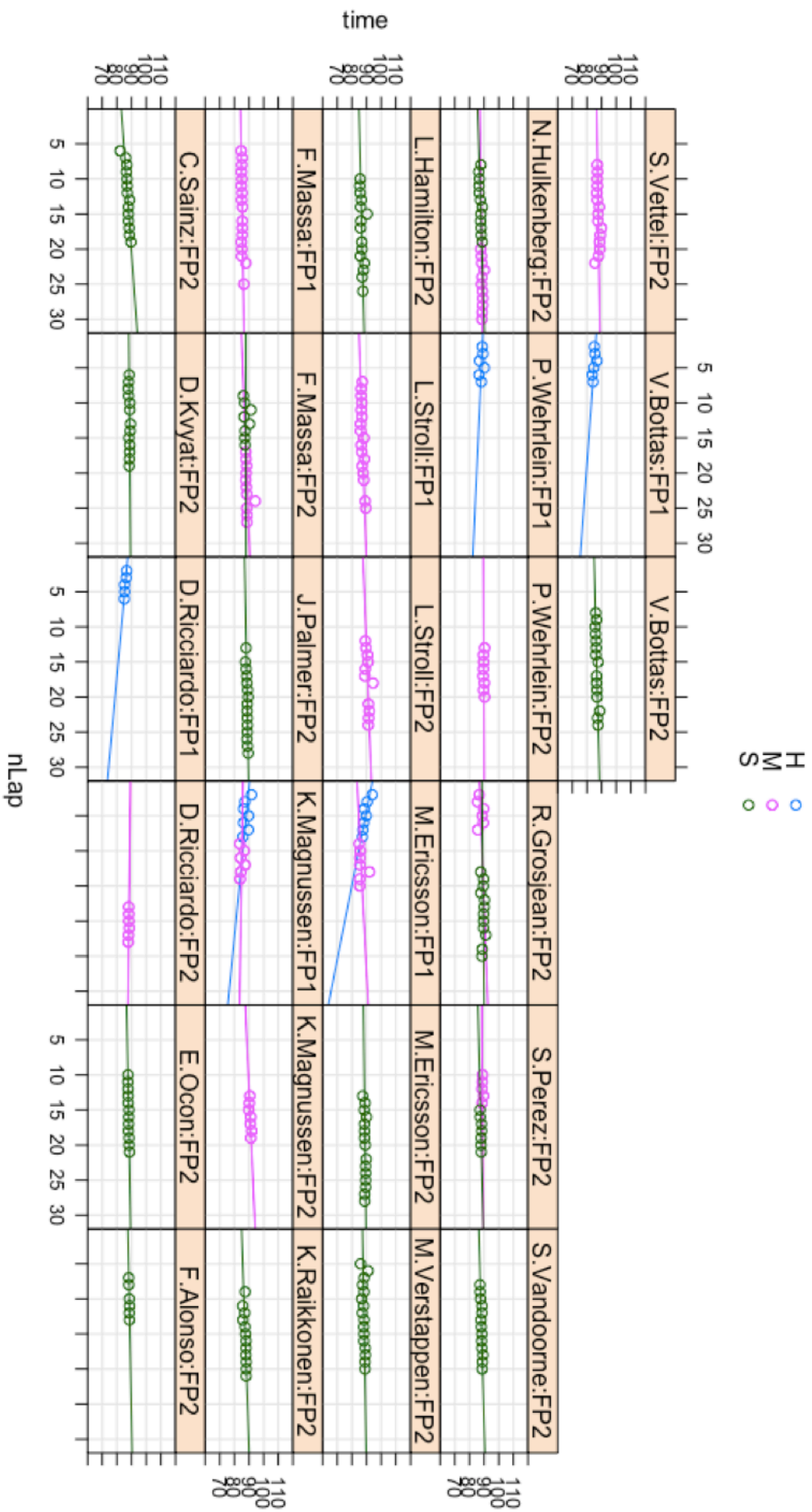
Normal Q-Q Plot

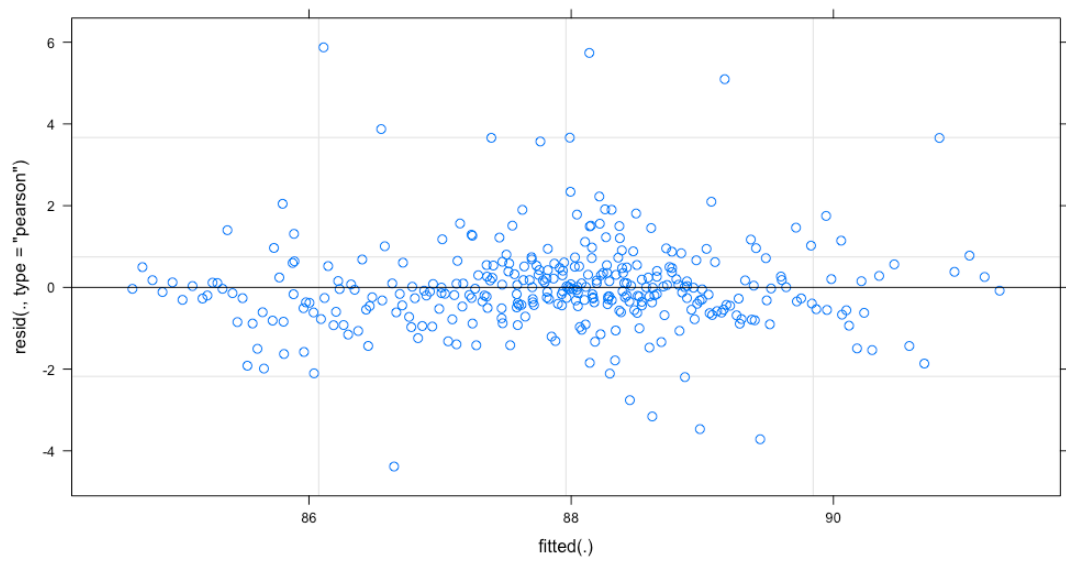


Histogram of resid(m1)

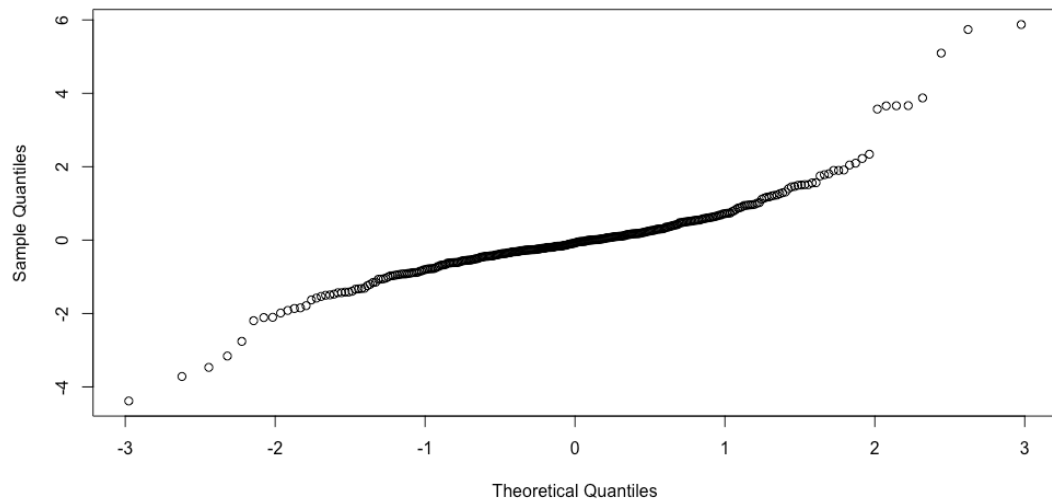




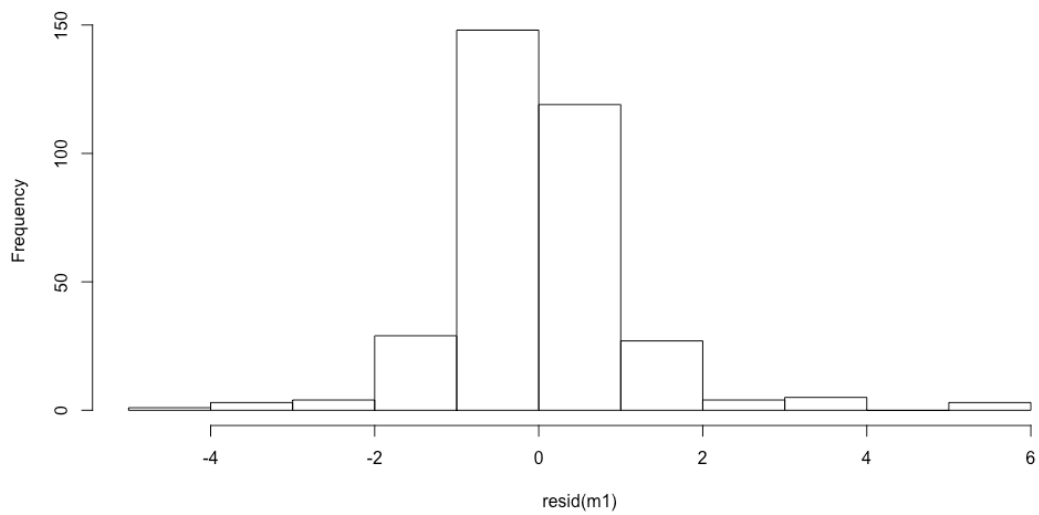


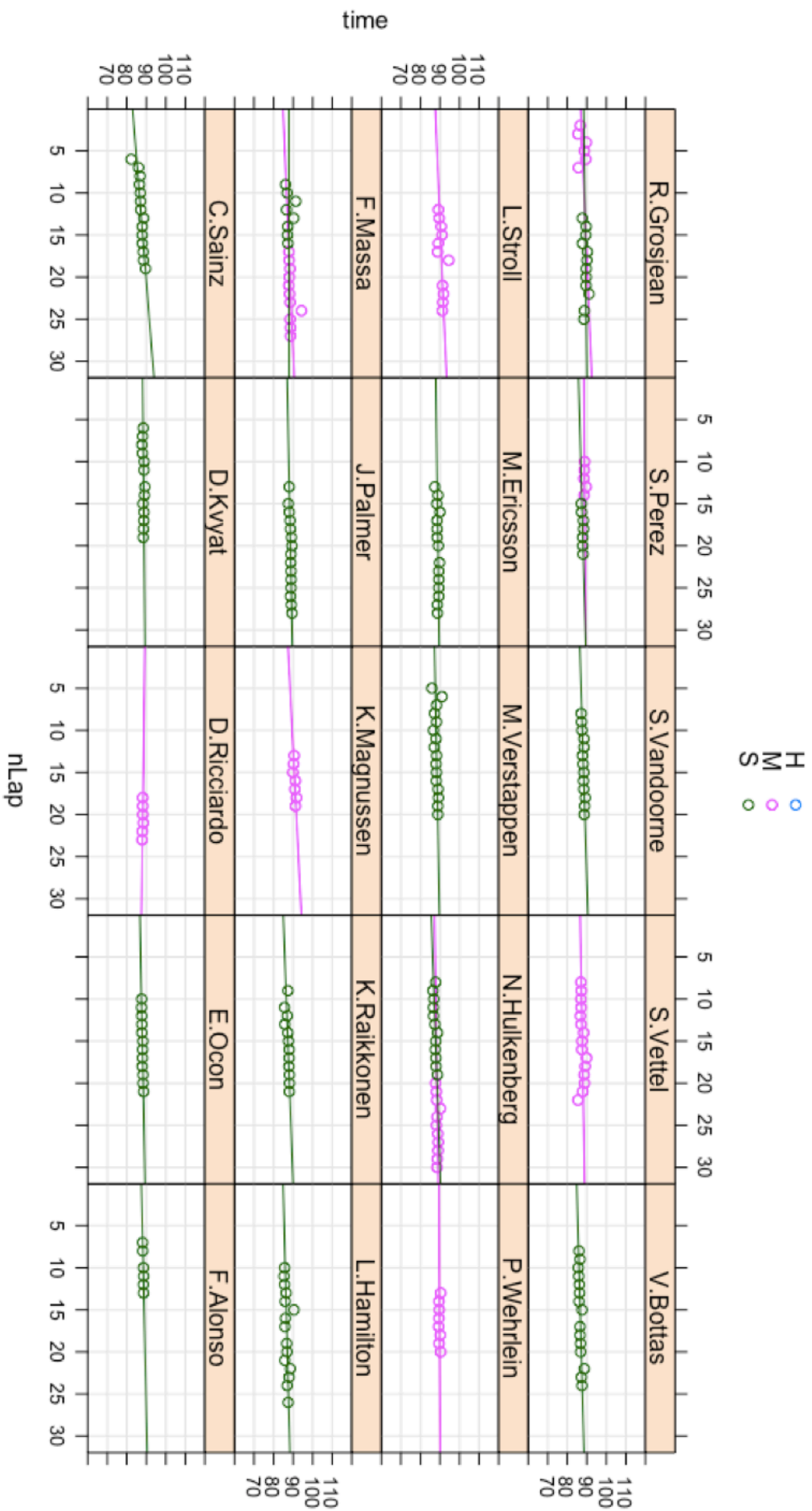


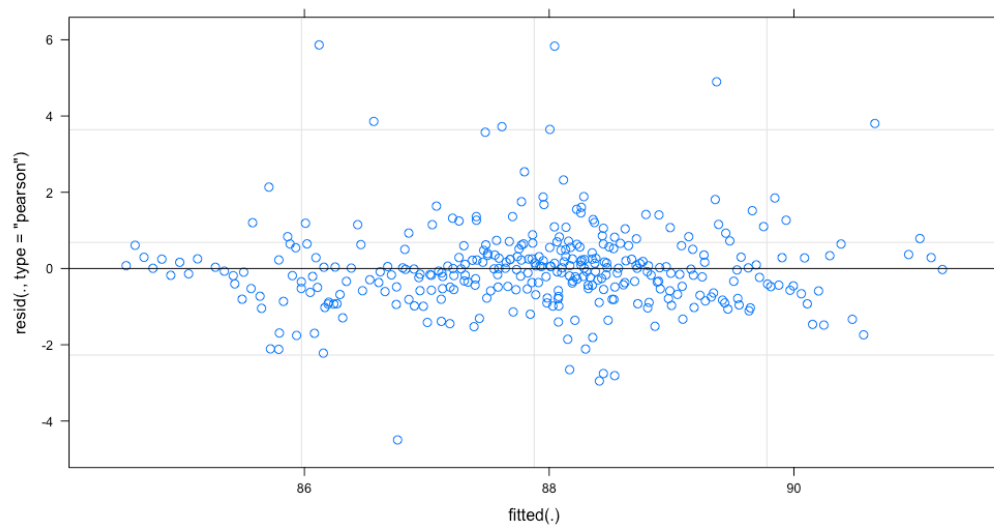
Normal Q-Q Plot



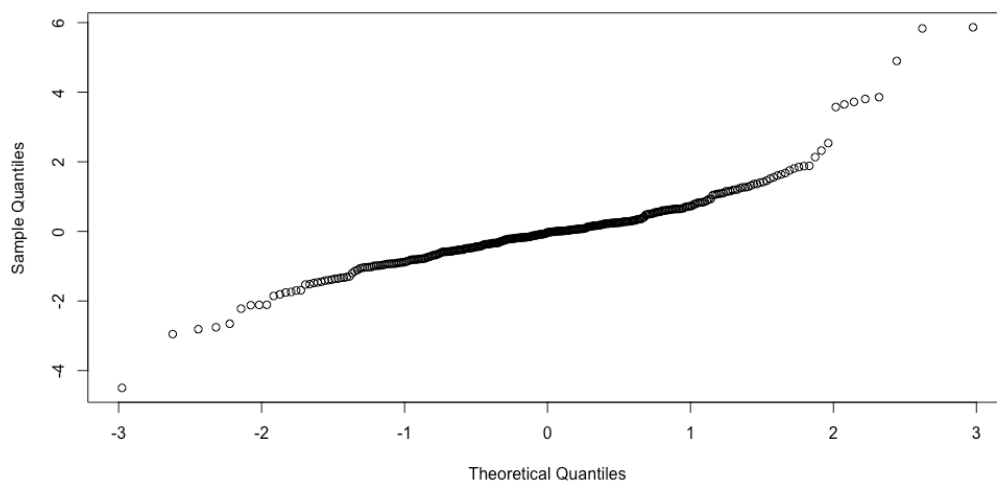
Histogram of resid(m1)



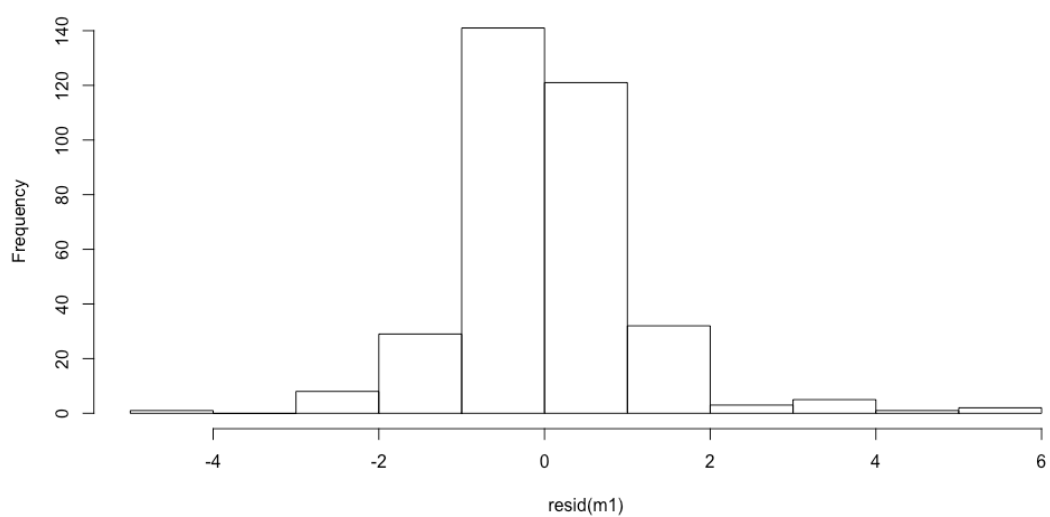


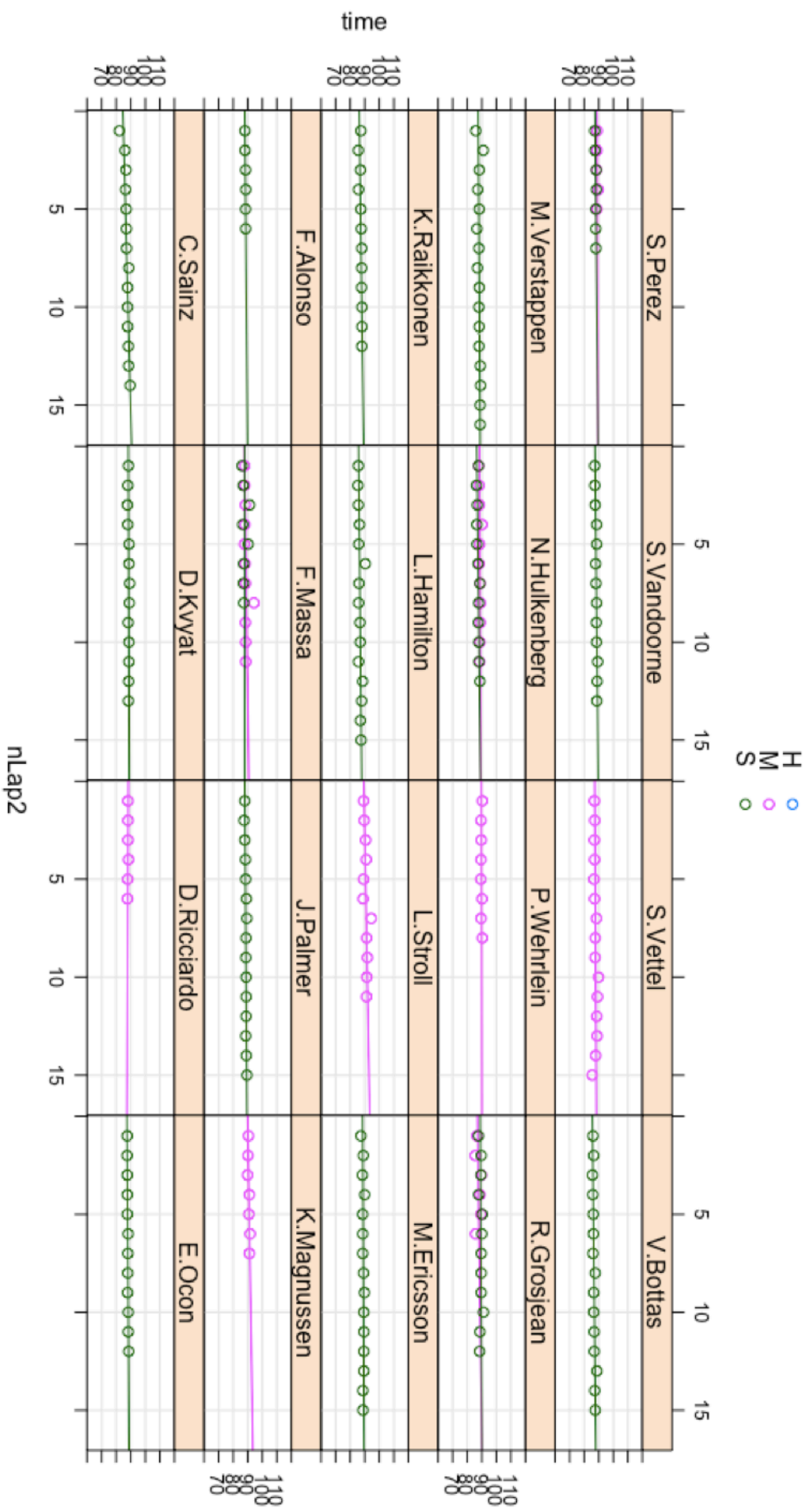


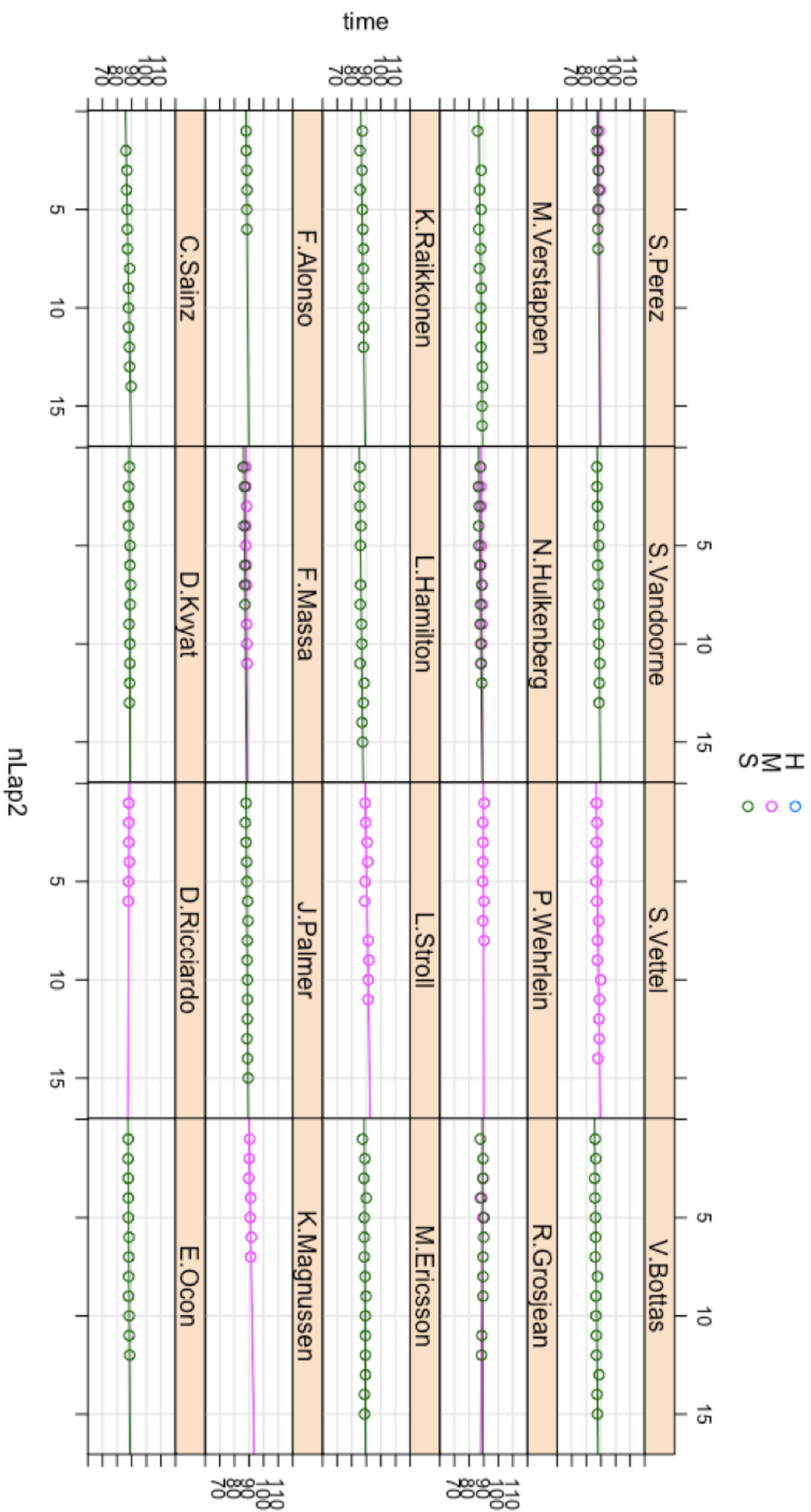
Normal Q-Q Plot

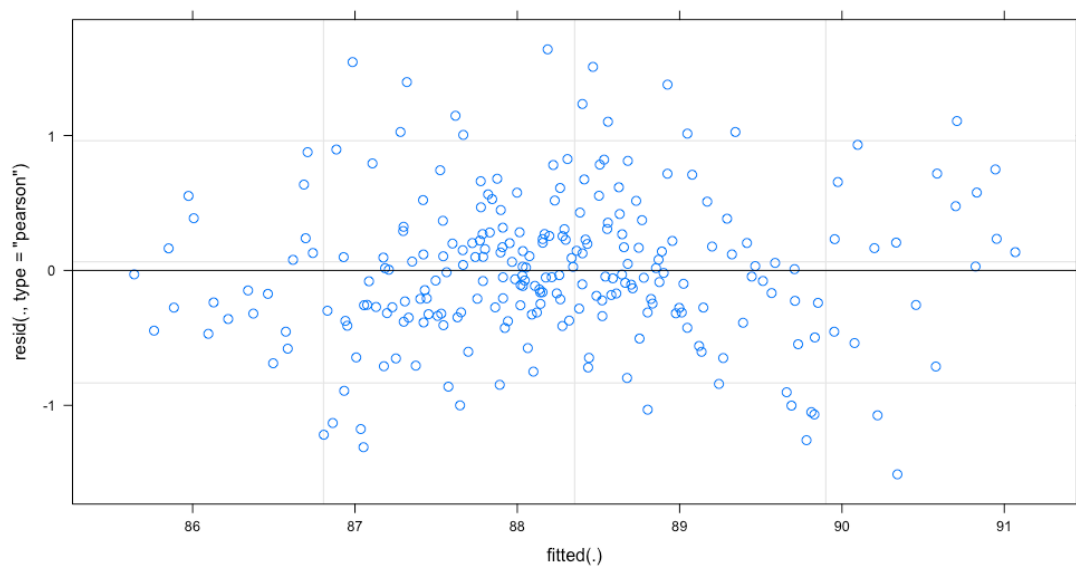


Histogram of resid(m1)

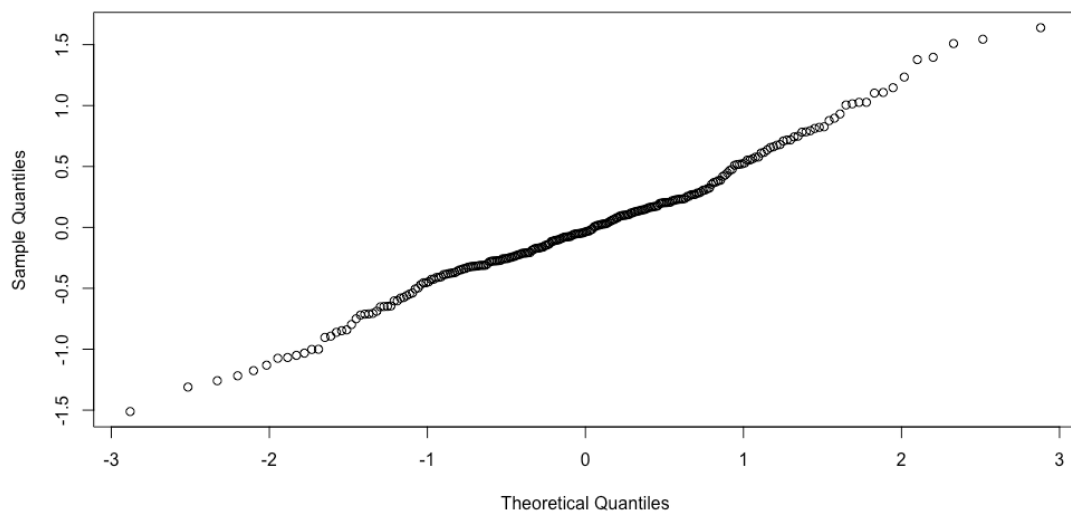




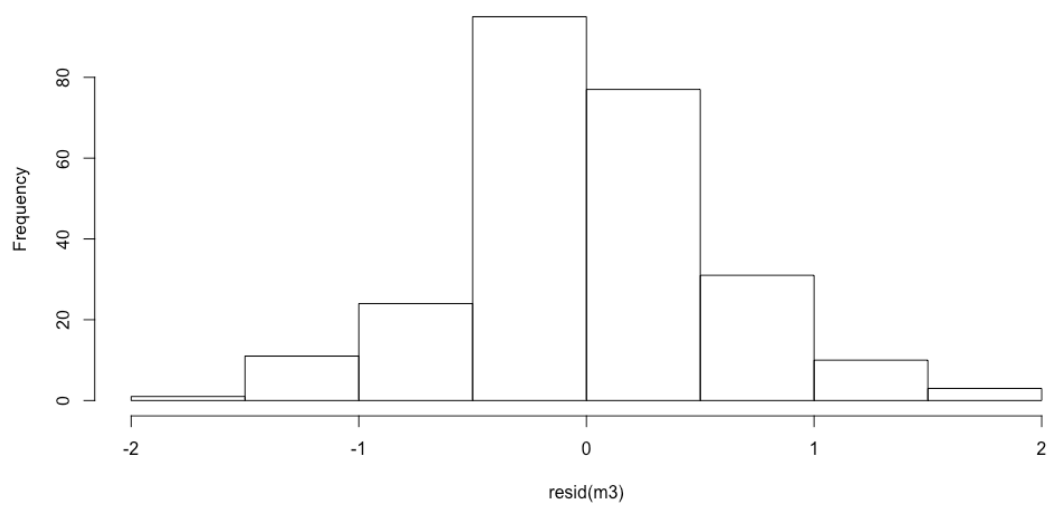




Normal Q-Q Plot



Histogram of resid(m3)



```
> summary(m3)
```

Linear mixed model fit by REML t-tests use Satterthwaite approximations to degrees of freedom [lmerMod]

Formula: time ~ nLap2 + Tyre + (1 | Team) + (1 | Driver)

Data: dadesFin

REML criterion at convergence: 503.9

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-2.67584	-0.55416	-0.06523	0.47521	2.90157

Random effects:

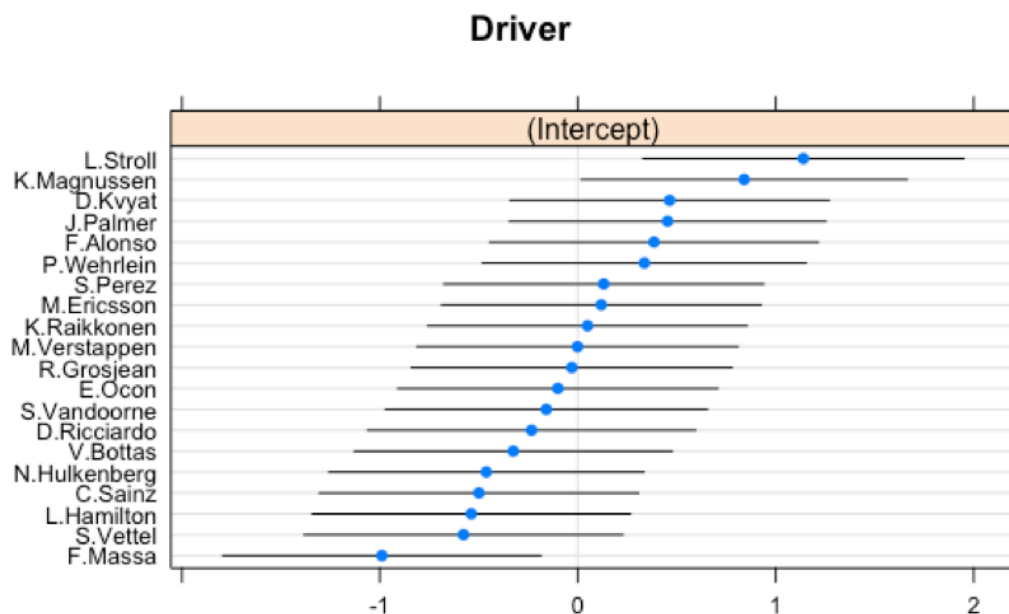
Groups	Name	Variance	Std.Dev.
Driver	(Intercept)	0.4275	0.6538
Team	(Intercept)	0.5840	0.7642
Residual		0.3191	0.5649

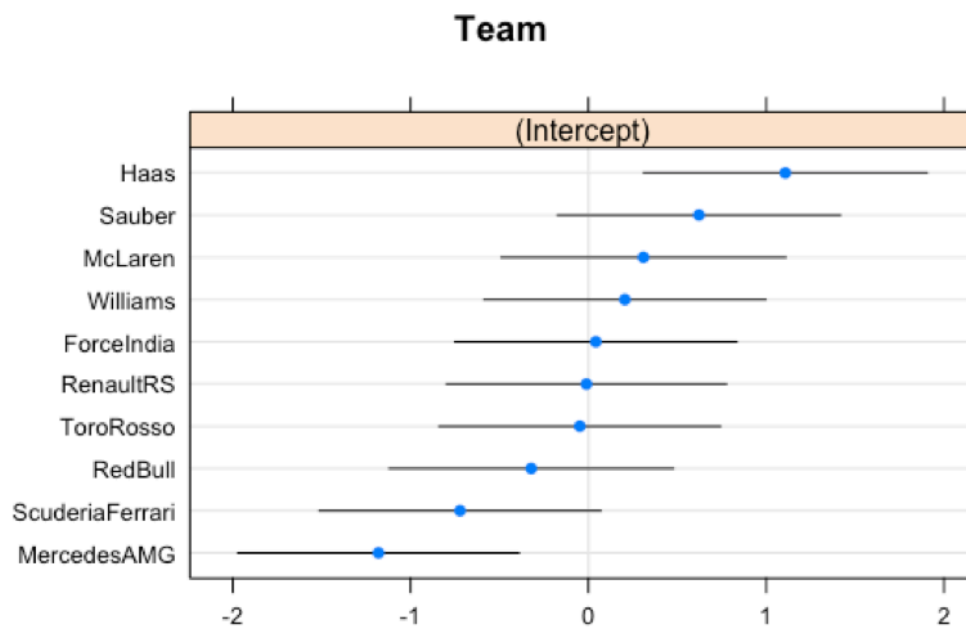
Number of obs: 252, groups: Driver, 20; Team, 10

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	88.263402	0.305556	11.770000	288.862	< 2e-16 ***
nLap2	0.122296	0.009849	233.180000	12.417	< 2e-16 ***
TyreS	-1.026723	0.140429	233.800000	-7.311	4.18e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



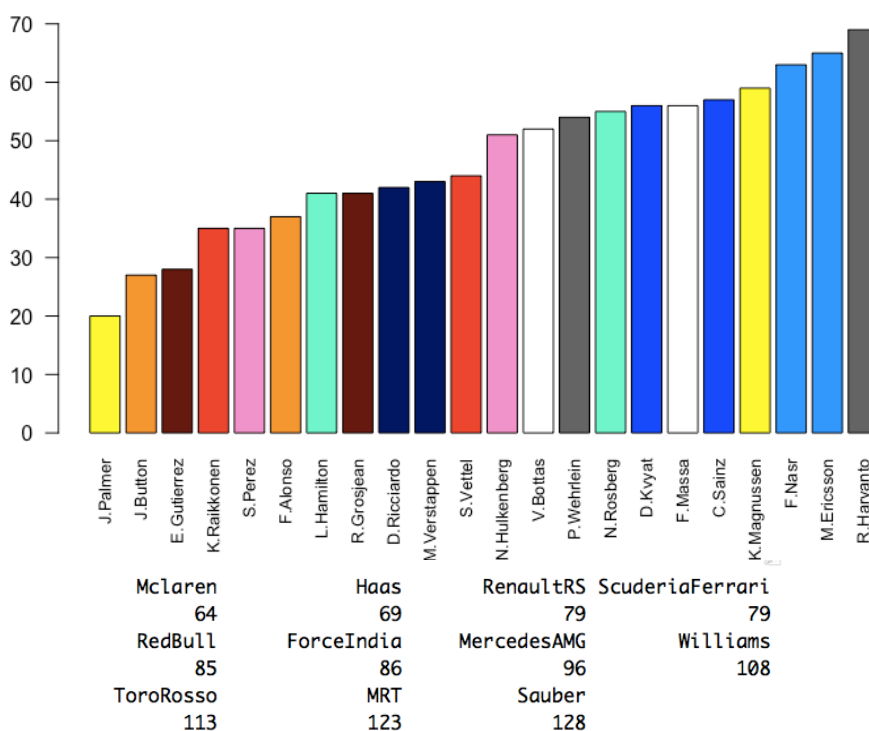


Annex B: GP 2016

Descriptiva univariada

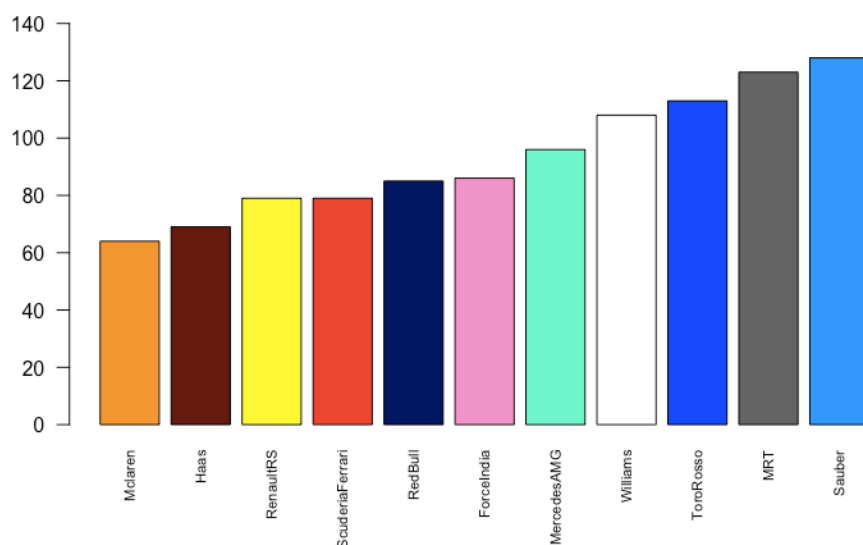
J. Palmer	J. Button	E. Gutierrez	K. Raikkonen	S. Perez
20	27	28	35	35
F. Alonso	L. Hamilton	R. Grosjean	D. Ricciardo	M. Verstappen
37	41	41	42	43
S. Vettel	N. Hulkenberg	V. Bottas	P. Wehrlein	N. Rosberg
44	51	52	54	55
D. Kvyat	F. Massa	C. Sainz	K. Magnussen	F. Nasr
56	56	57	59	63
M. Ericsson	R. Haryanto			
65	69			

Voltes per pilot GP 2016



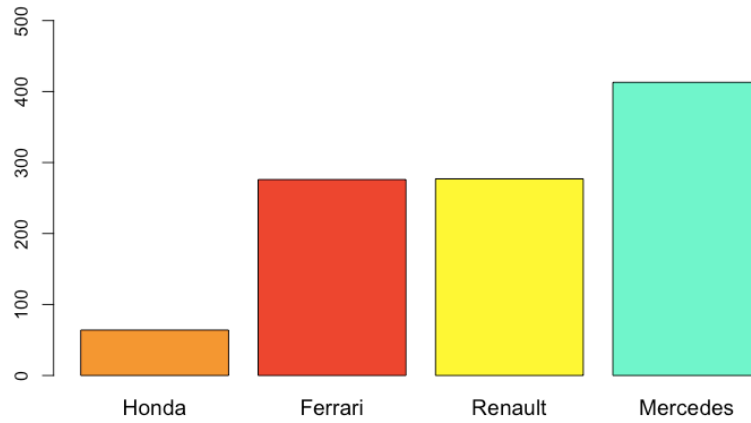
McLaren	Haas	RenaultRS	ScuderiaFerrari
64	69	79	79
RedBull	ForceIndia	MercedesAMG	Williams
85	86	96	108
ToroRosso	MRT	Sauber	
113	123	128	

Voltes per equips GP 2016



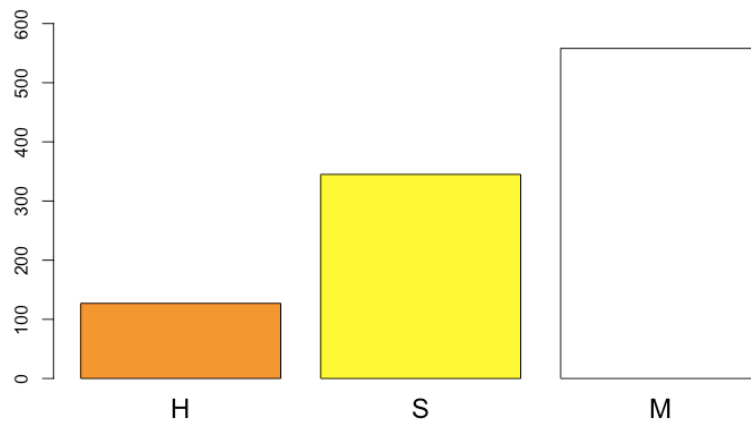
Honda	Ferrari	Renault	Mercedes
64	276	277	413

Voltes per motor GP 2016



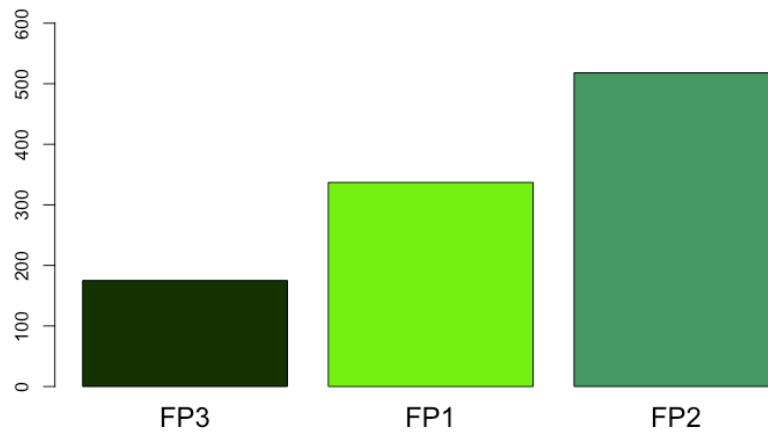
H	S	M
127	345	558

Voltes per pneumàtic GP 2016



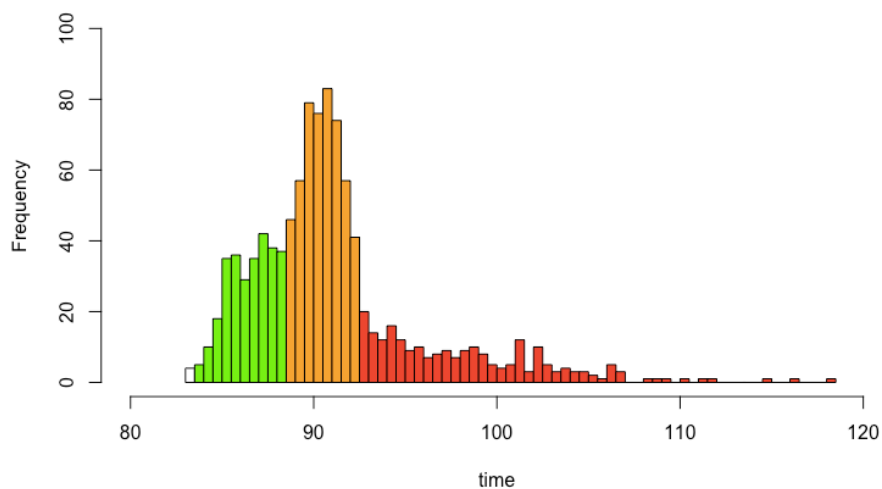
FP3	FP1	FP2
175	337	518

Voltes per sessió GP 2016

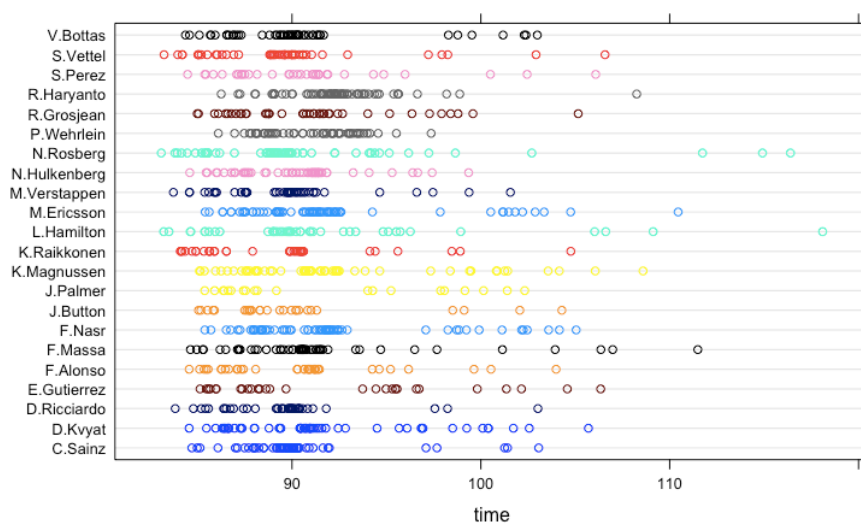


Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
83.08	88.04	90.31	91.08	92.17	118.10

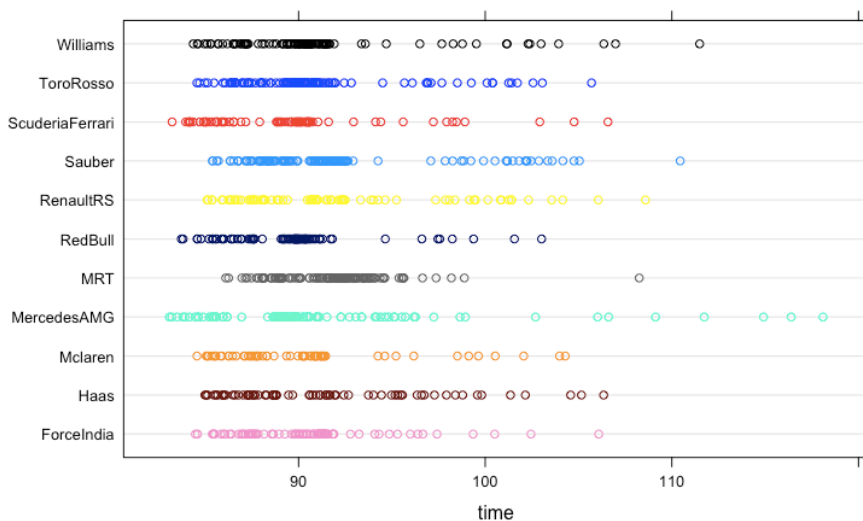
Frequència dels temps GP 2016



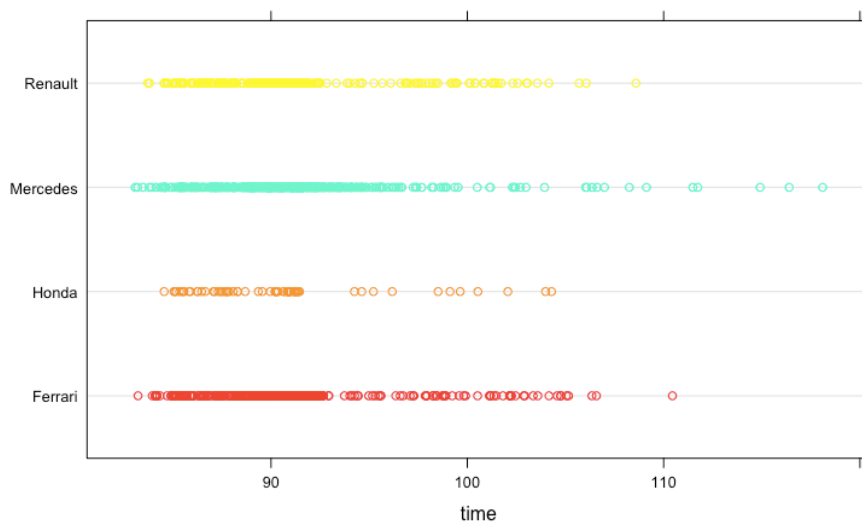
Driver vs time - GP 2016



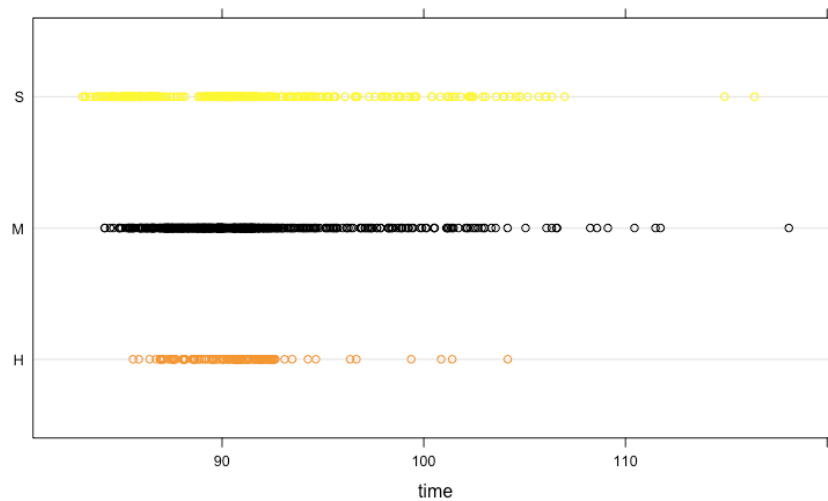
Team vs time - GP 2016



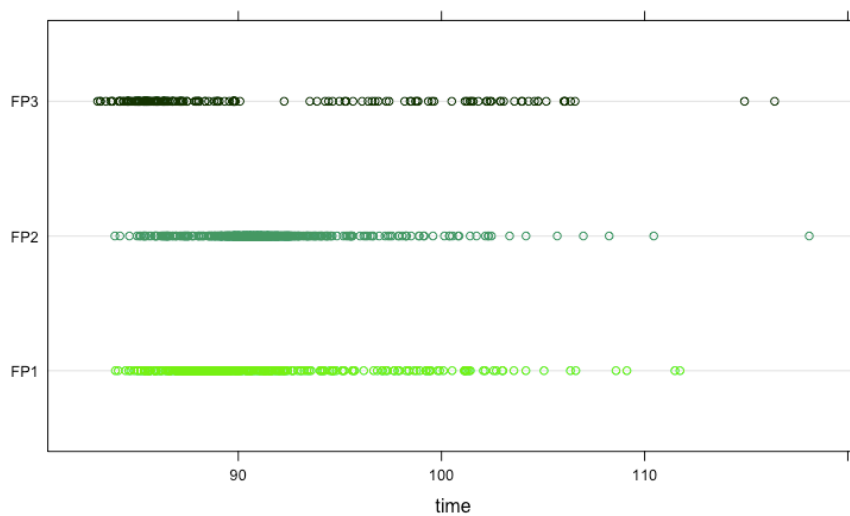
Engine vs time - GP 2016



Tyre vs time - GP 2016

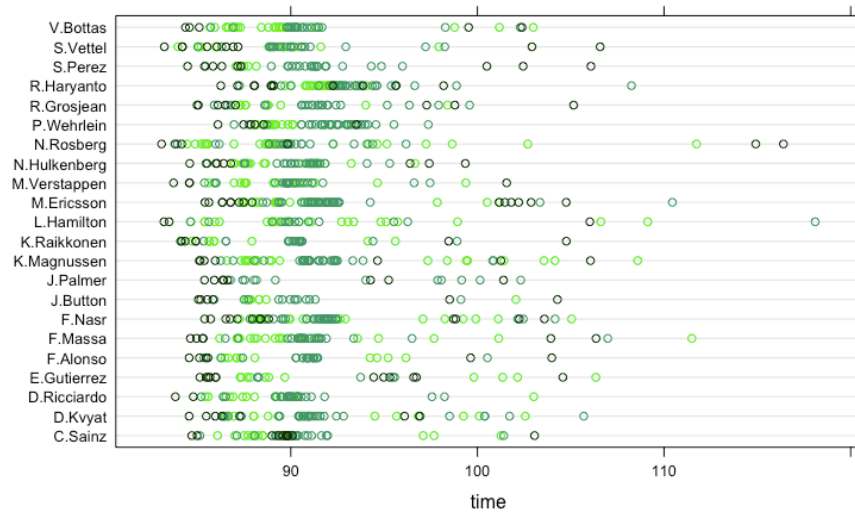


Session vs time - GP 2016

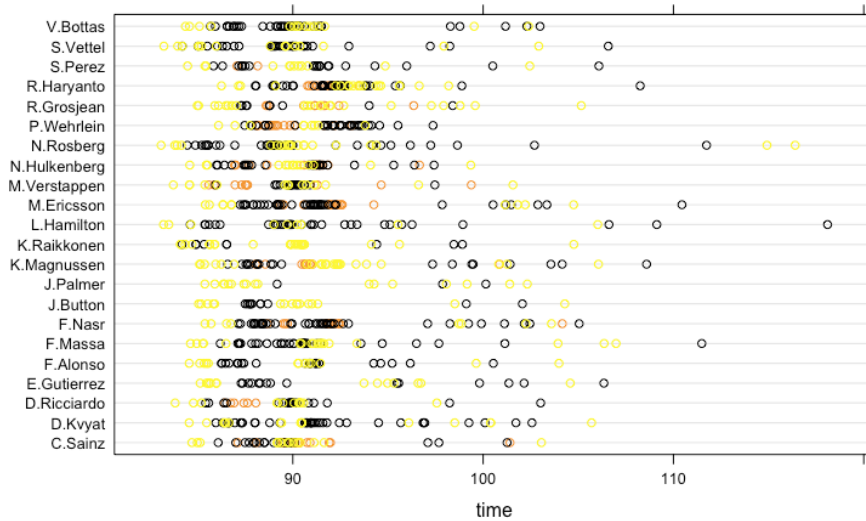


Descriptiva bivariada

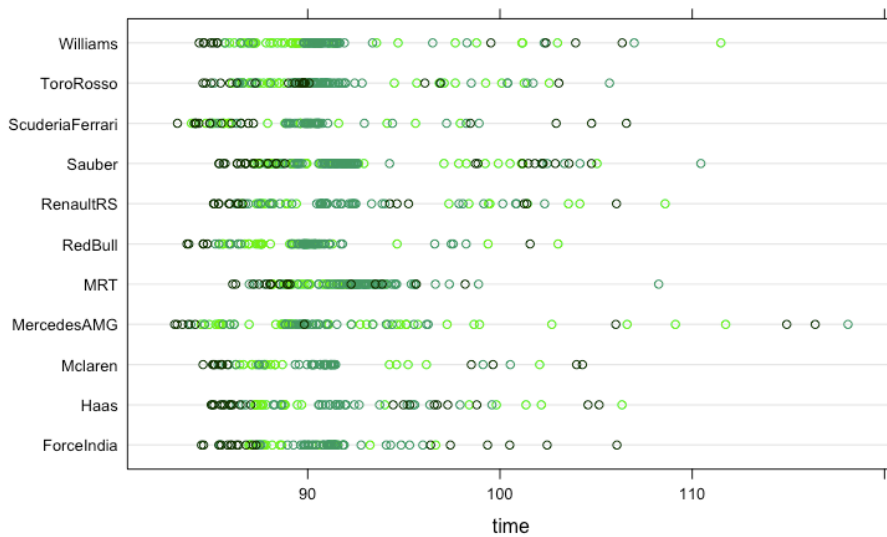
Driver vs time segons Session - GP 2016



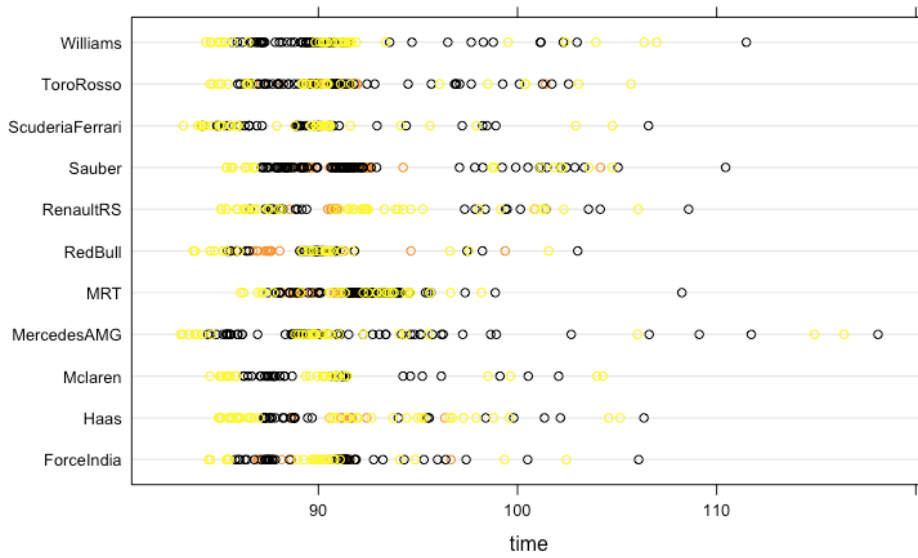
Driver vs time segons Tyre - GP 2016



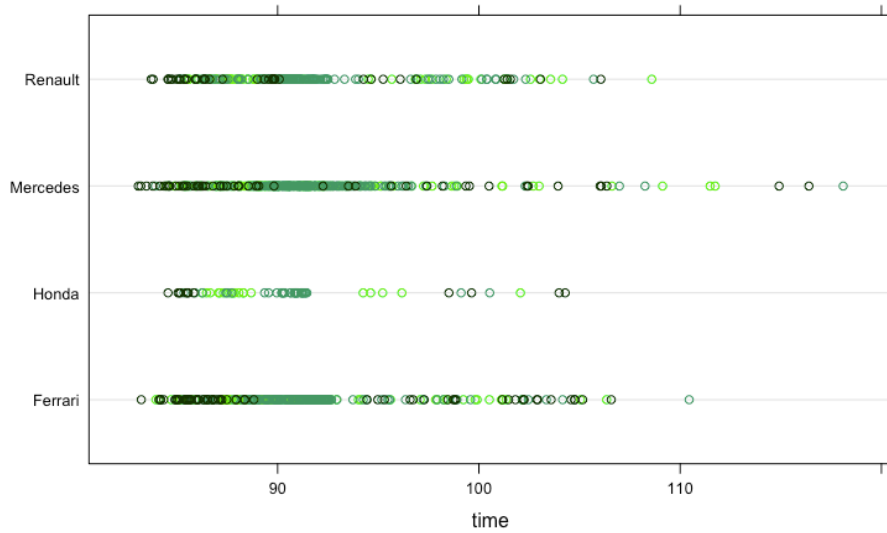
Team vs time segons Session - GP 2016



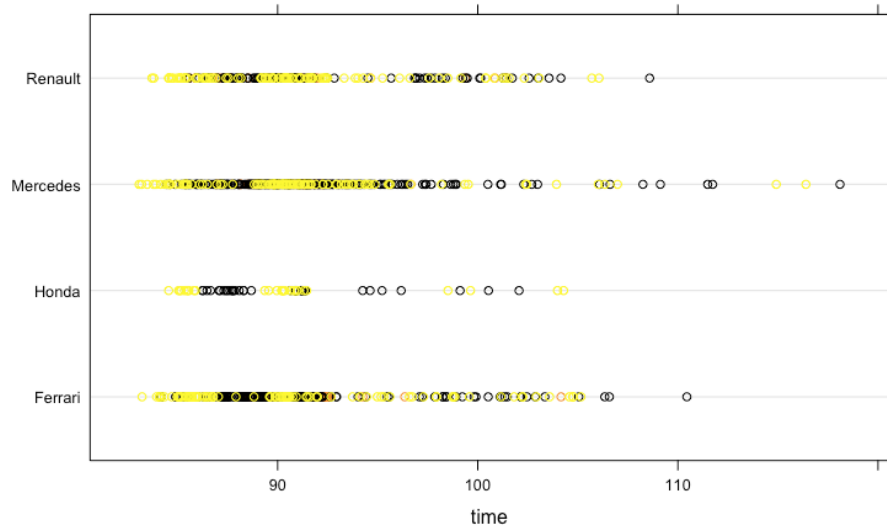
Team vs time segons Tyre - GP 2016



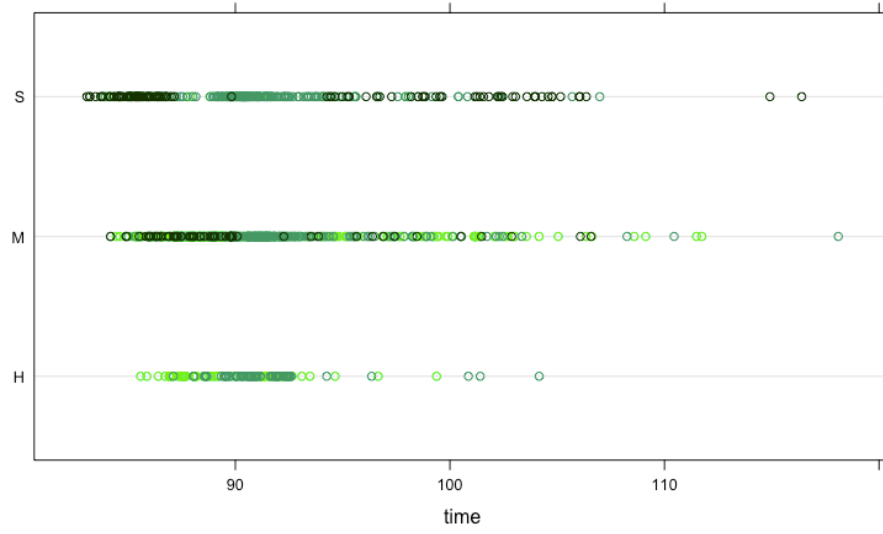
Engine vs time segons Session - GP 2016



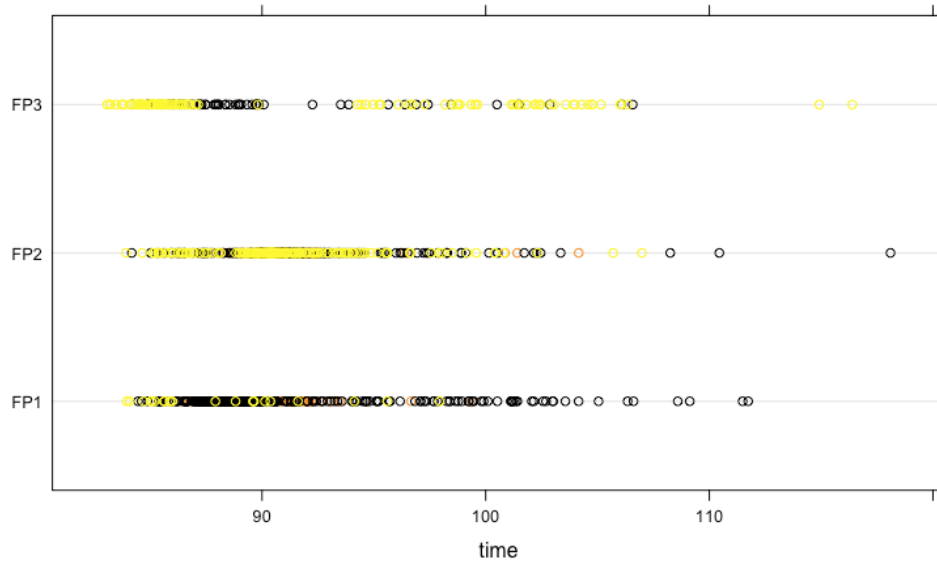
Engine vs time segons Tyre - GP 2016



Tyre vs time segons Session - GP 2016

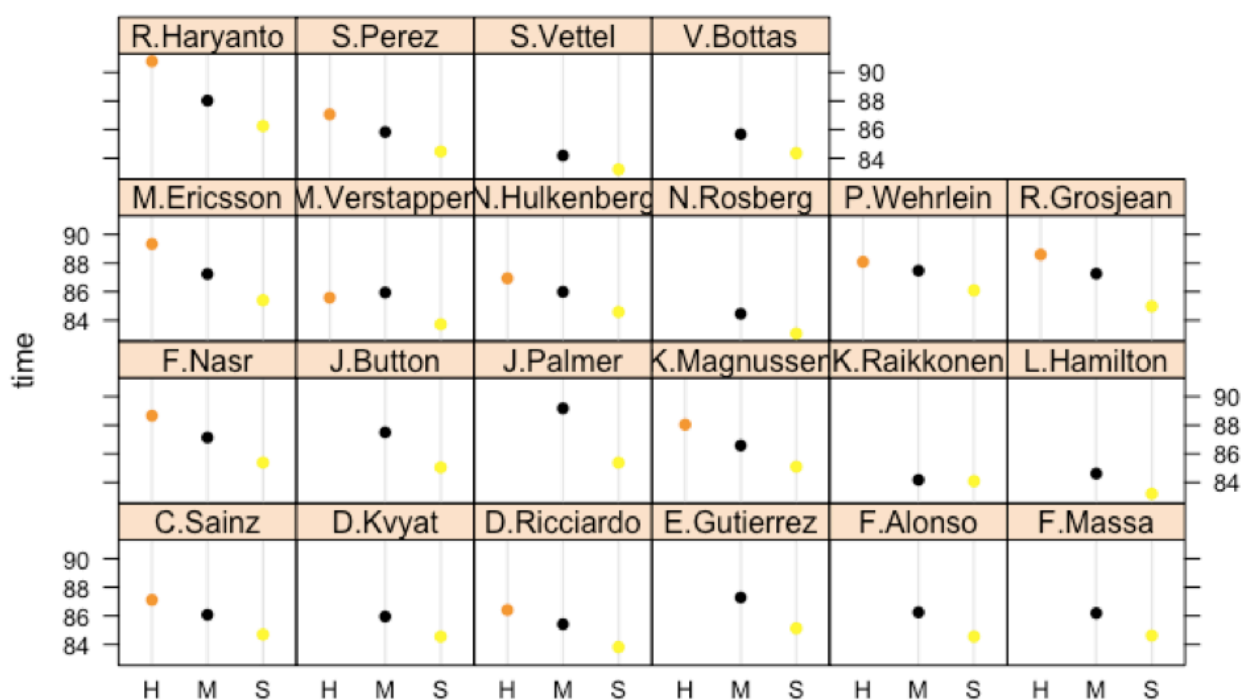


Session vs time segons Tyre - GP 2016

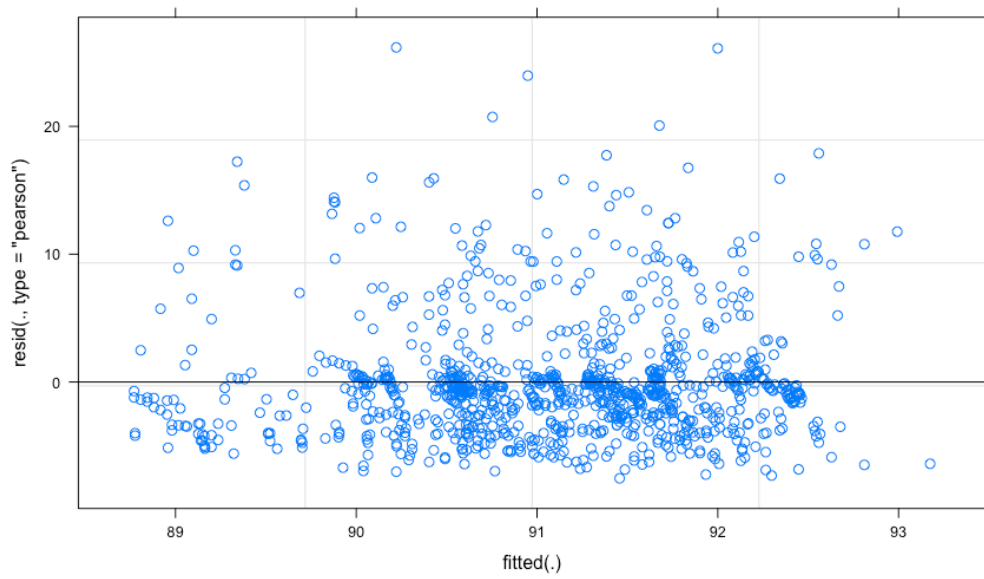


Resultats

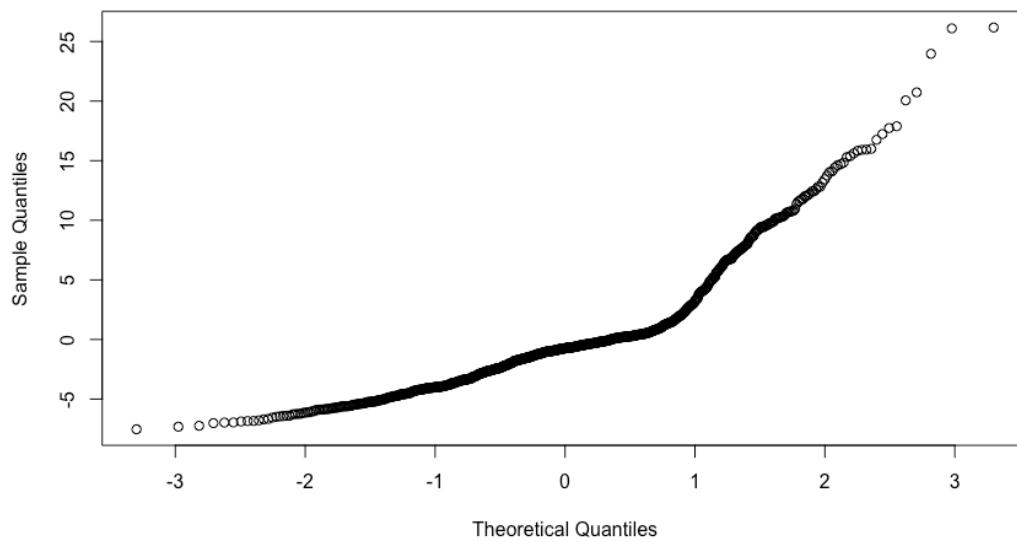
GridPosition	Number	Driver	Team	Engine	Tyre	Session	time
1	6	N.Rosberg	MercedesAMG	Mercedes	S	FP3	83.078
2	44	L.Hamilton	MercedesAMG	Mercedes	S	FP3	83.204
3	5	S.Vettel	ScuderiaFerrari	Ferrari	S	FP3	83.225
4	33	M.Verstappen	RedBull	Renault	S	FP3	83.719
5	3	D.Ricciardo	RedBull	Renault	S	FP3	83.816
6	7	K.Raikkonen	ScuderiaFerrari	Ferrari	S	FP1	84.089
7	77	V.Bottas	Williams	Mercedes	S	FP3	84.356
8	11	S.Perez	ForceIndia	Mercedes	S	FP3	84.472
9	26	D.Kvyat	ToroRosso	Renault	S	FP3	84.553
10	14	F.Alonso	Mclaren	Honda	S	FP3	84.555
11	27	N.Hulkenberg	ForceIndia	Mercedes	S	FP3	84.585
12	19	F.Massa	Williams	Mercedes	S	FP3	84.621
13	55	C.Sainz	ToroRosso	Renault	S	FP3	84.695
14	8	R.Grosjean	Haas	Ferrari	S	FP3	84.981
15	22	J.Button	Mclaren	Honda	S	FP3	85.051
16	20	K.Magnussen	RenaultRS	Renault	S	FP3	85.100
17	21	E.Gutierrez	Haas	Ferrari	S	FP3	85.130
18	30	J.Palmer	RenaultRS	Renault	S	FP3	85.376
19	12	F.Nasr	Sauber	Ferrari	S	FP3	85.383
20	9	M.Ericsson	Sauber	Ferrari	S	FP3	85.401
21	94	P.Wehrlein	MRT	Mercedes	S	FP3	86.097
22	88	R.Haryanto	MRT	Mercedes	S	FP3	86.251



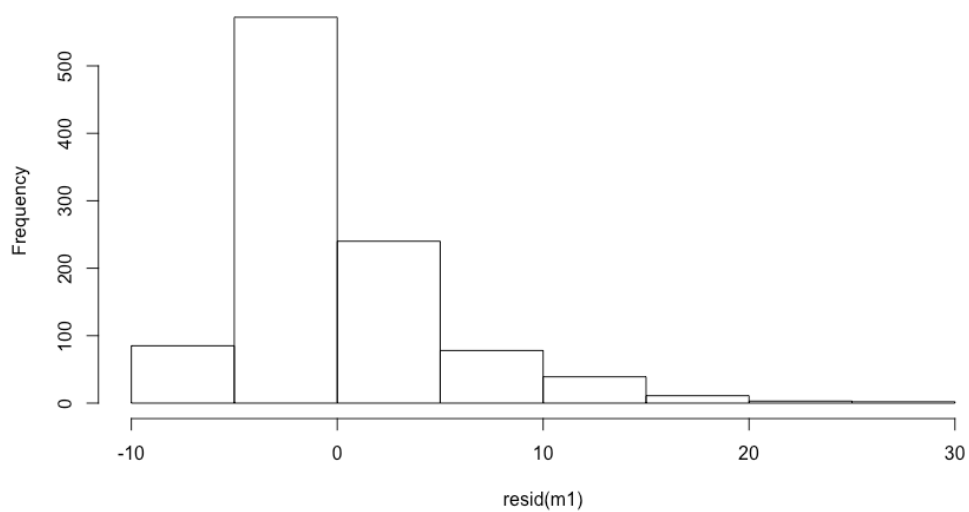
GridPosition	BesttimeSoft	BesttimeMedium	BesttimeHard
1	N.Rosberg	K.Raikkonen	M.Verstappen
2	L.Hamilton	S.Vettel	D.Ricciardo
3	S.Vettel	N.Rosberg	N.Hulkenberg
4	M.Verstappen	L.Hamilton	S.Perez
5	D.Ricciardo	D.Ricciardo	C.Sainz
6	K.Raikkonen	V.Bottas	K.Magnussen
7	V.Bottas	S.Perez	P.Wehrlein
8	S.Perez	M.Verstappen	R.Grosjean
9	D.Kvyat	D.Kvyat	F.Nasr
10	F.Alonso	N.Hulkenberg	M.Ericsson
11	N.Hulkenberg	C.Sainz	R.Haryanto
12	F.Massa	F.Massa	<NA>
13	C.Sainz	F.Alonso	<NA>
14	R.Grosjean	K.Magnussen	<NA>
15	J.Button	F.Nasr	<NA>
16	K.Magnussen	M.Ericsson	<NA>
17	E.Gutierrez	R.Grosjean	<NA>
18	J.Palmer	E.Gutierrez	<NA>
19	F.Nasr	P.Wehrlein	<NA>
20	M.Ericsson	J.Button	<NA>
21	P.Wehrlein	R.Haryanto	<NA>
22	R.Haryanto	J.Palmer	<NA>



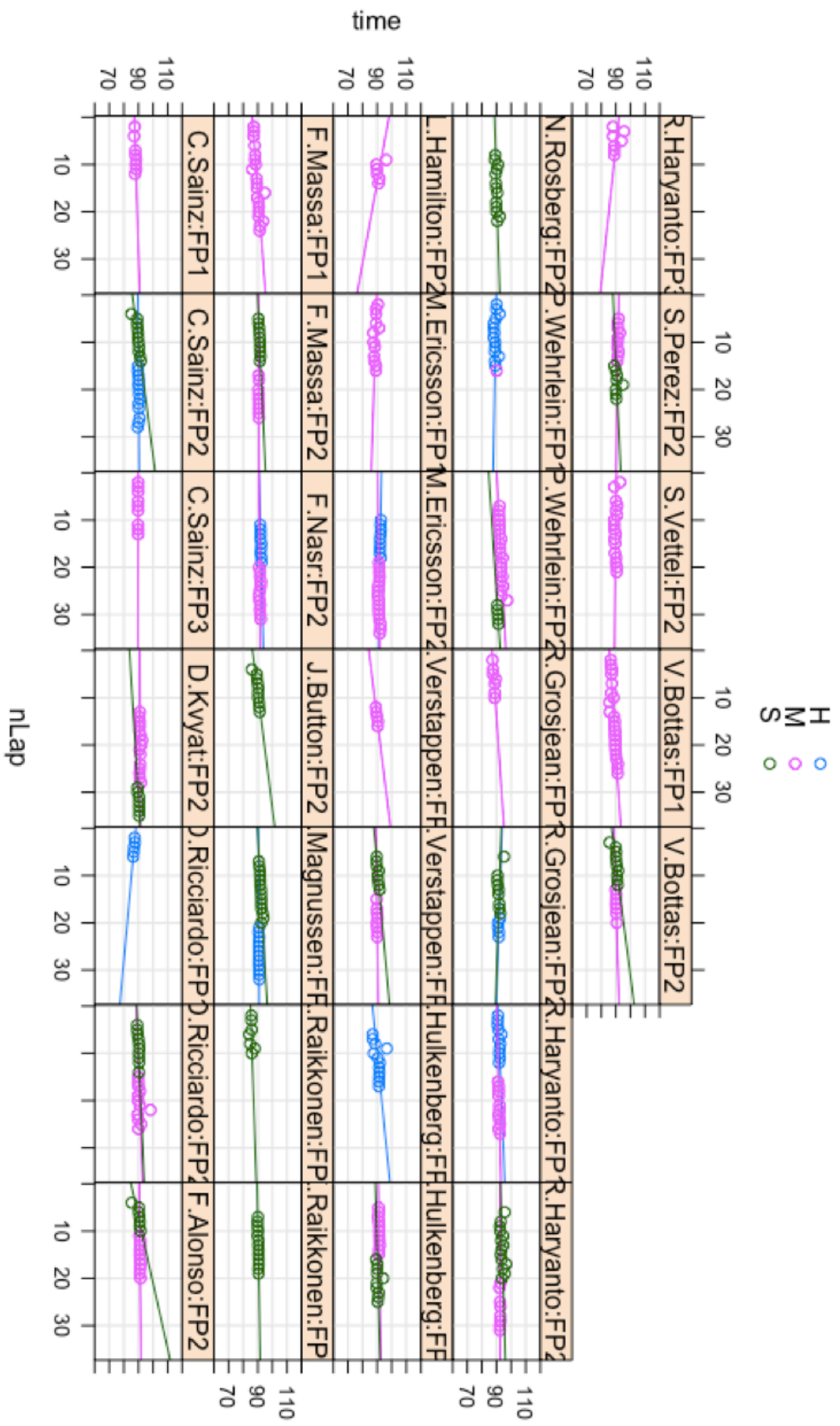
Normal Q-Q Plot

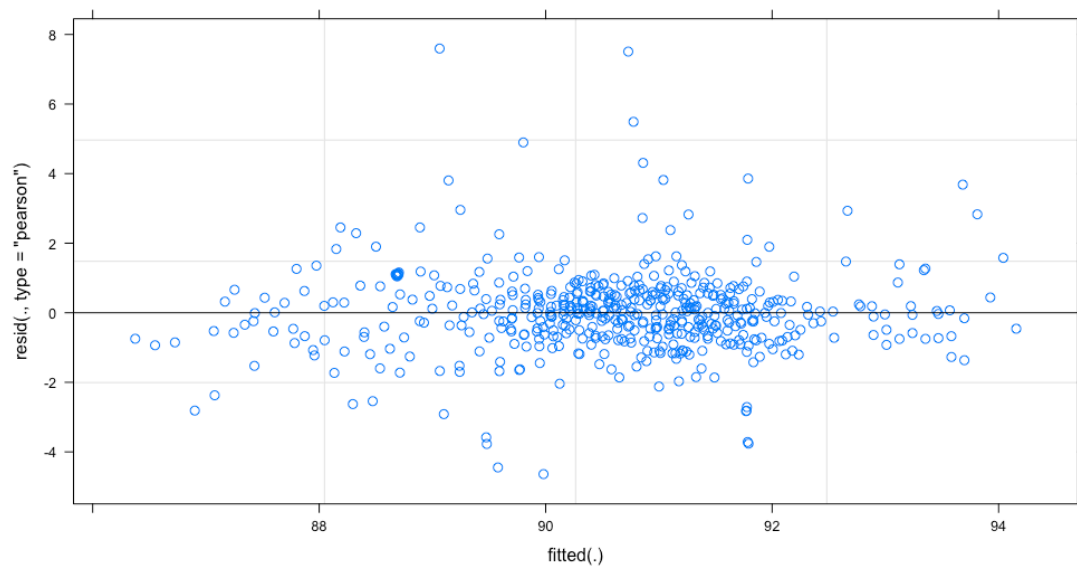


Histogram of resid(m1)

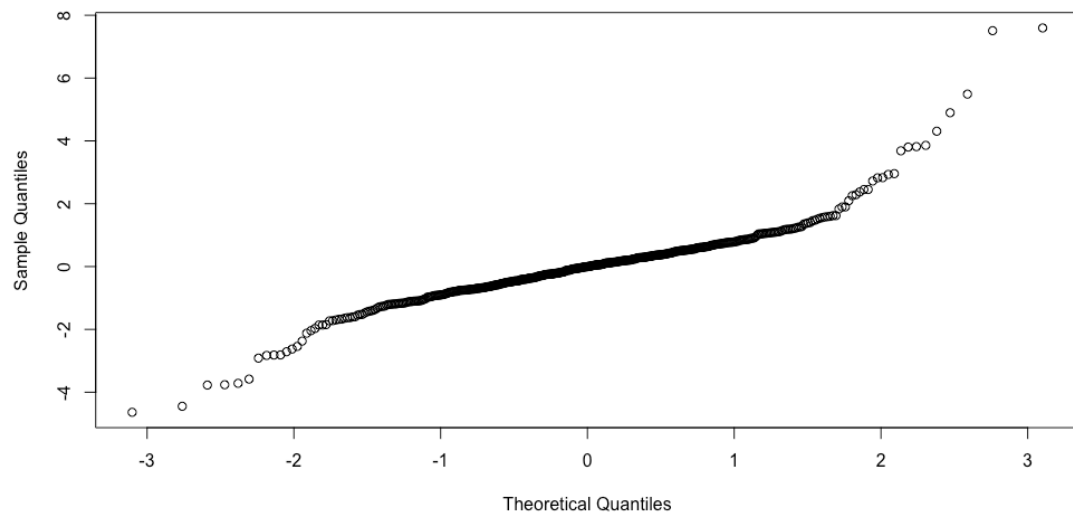




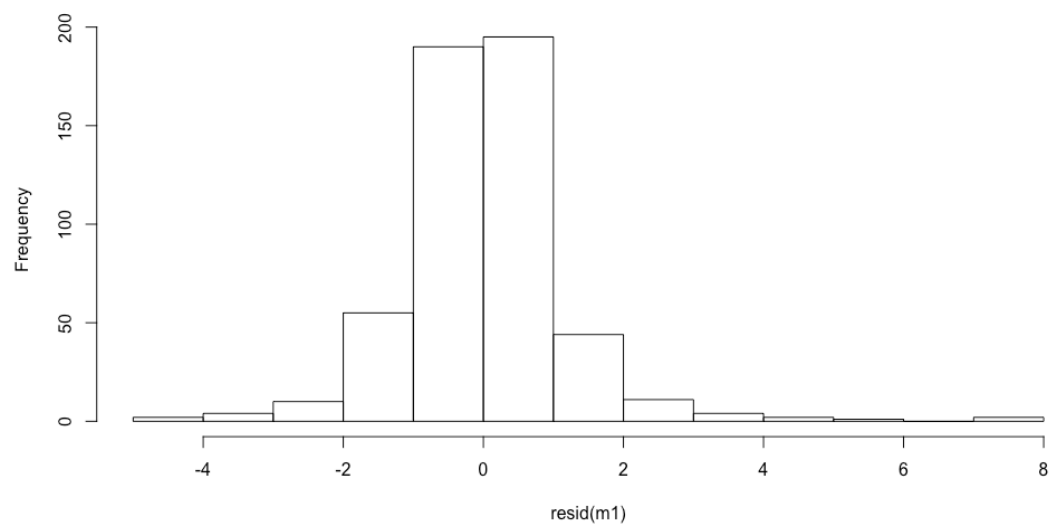


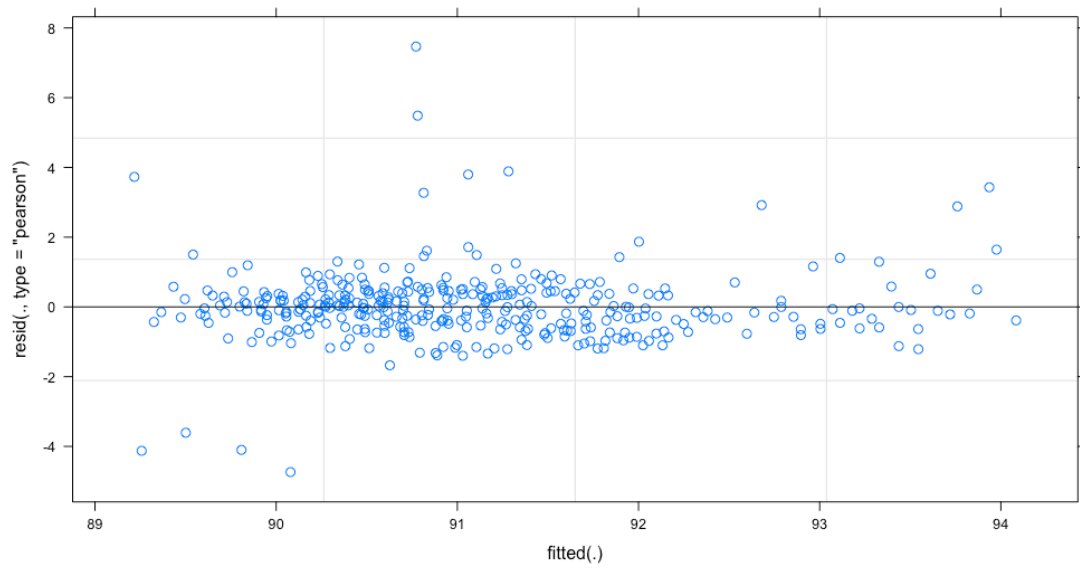


Normal Q-Q Plot

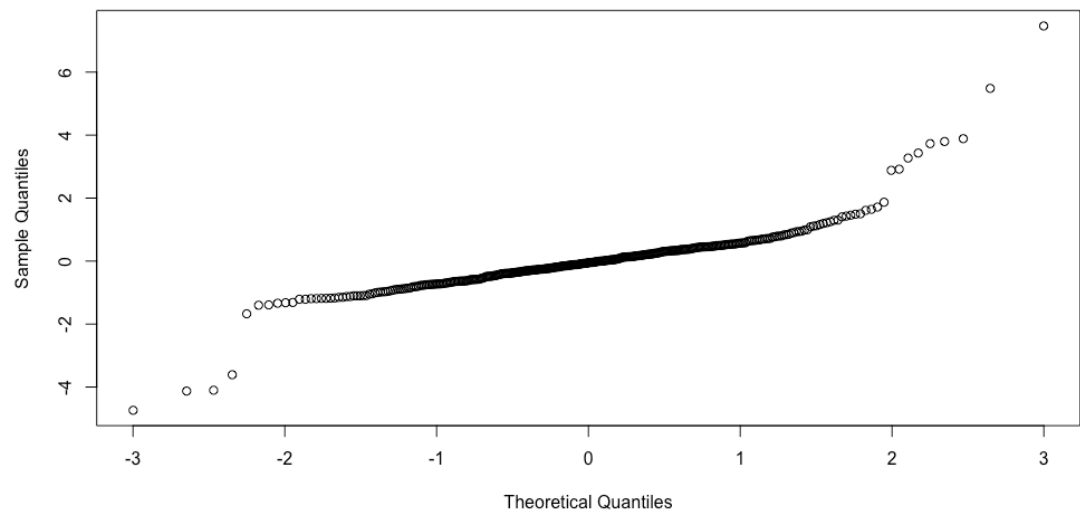


Histogram of resid(m1)

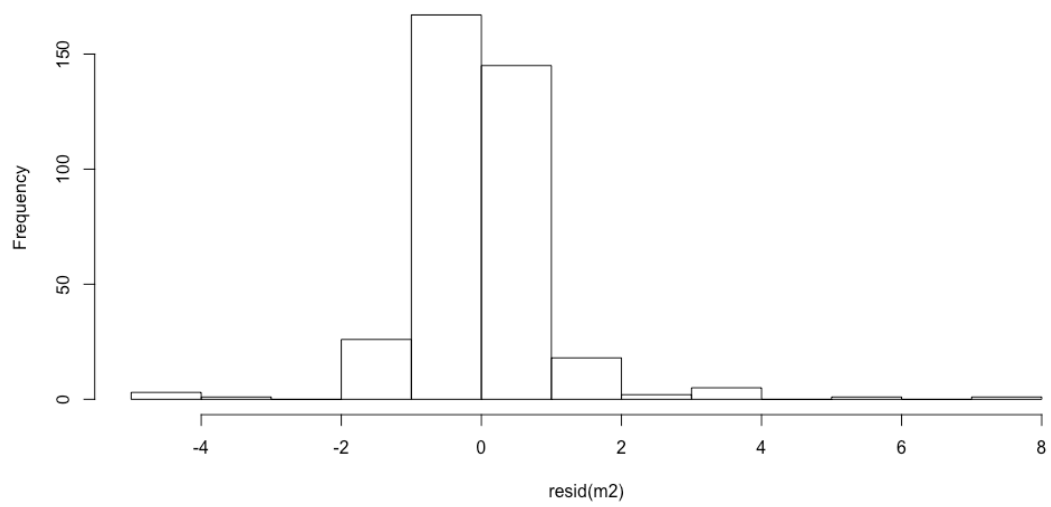


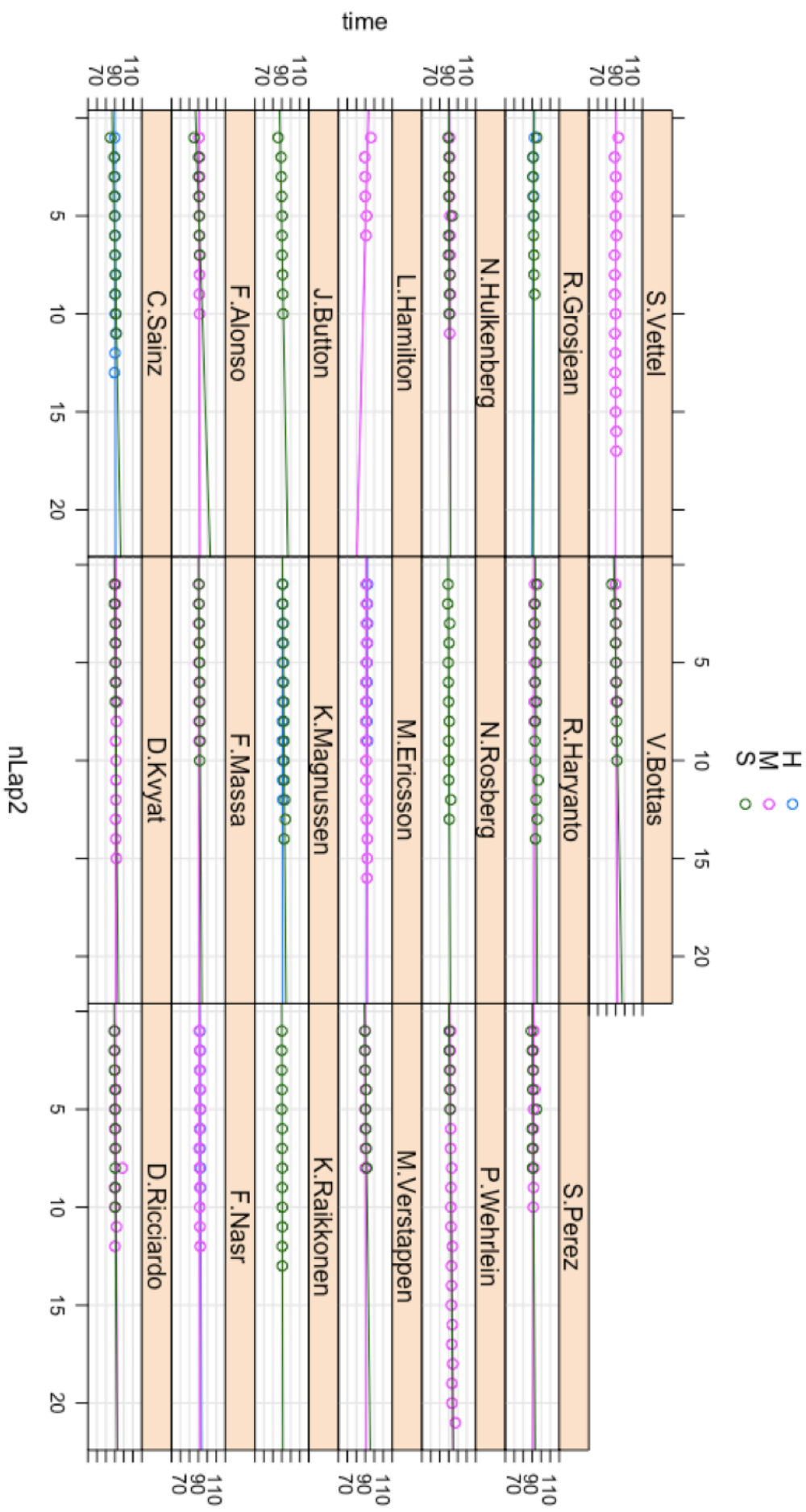


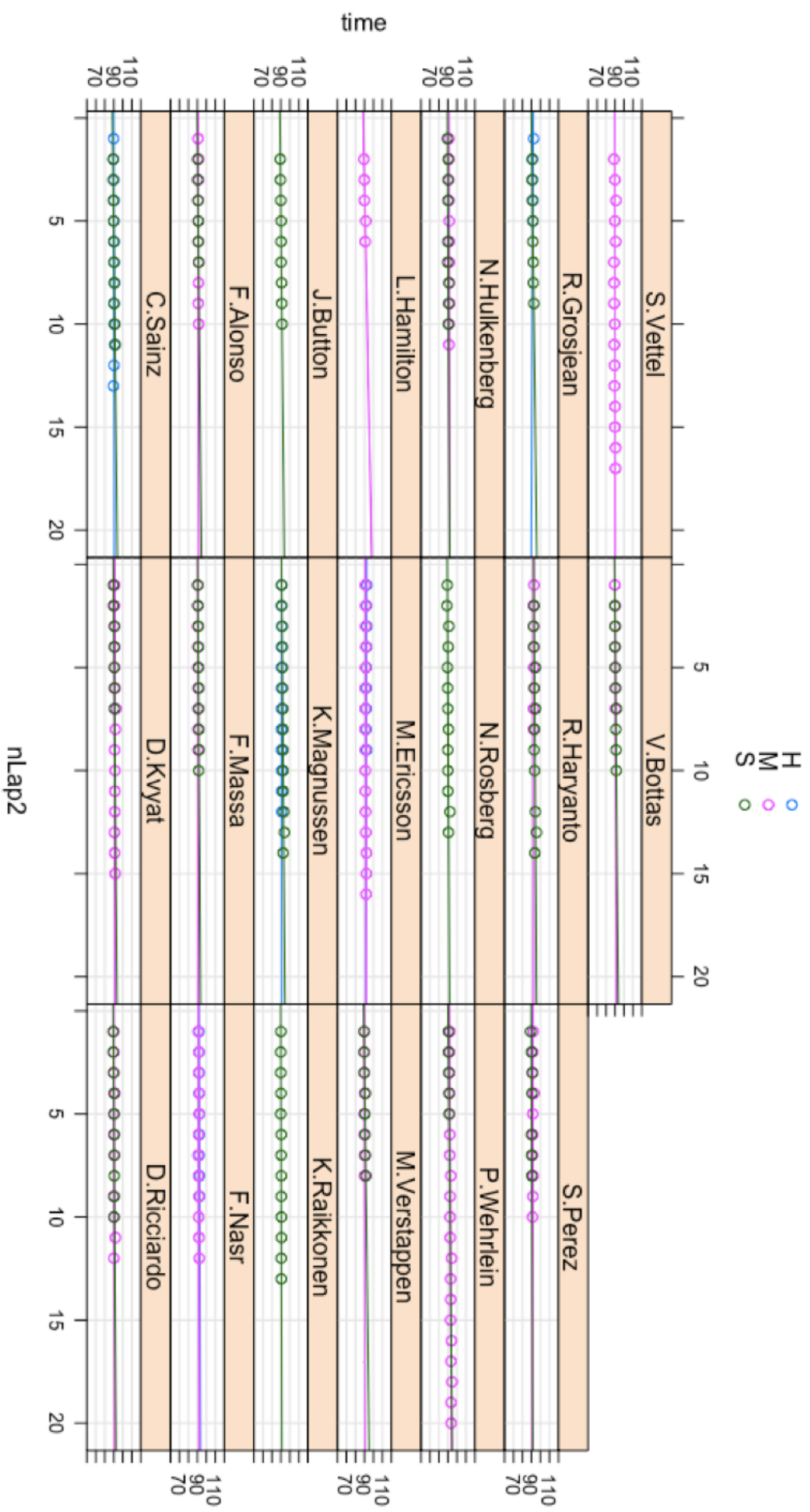
Normal Q-Q Plot

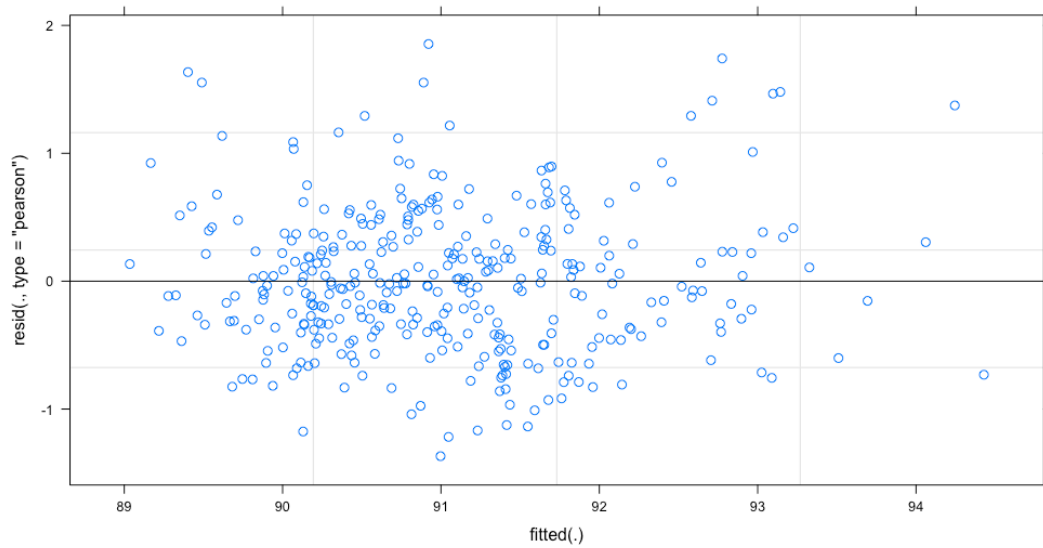


Histogram of resid(m2)

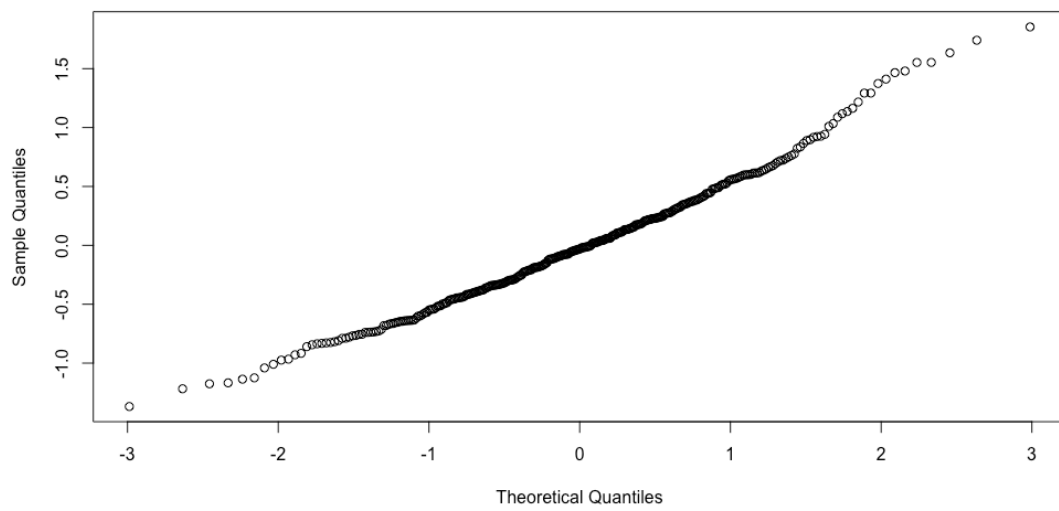




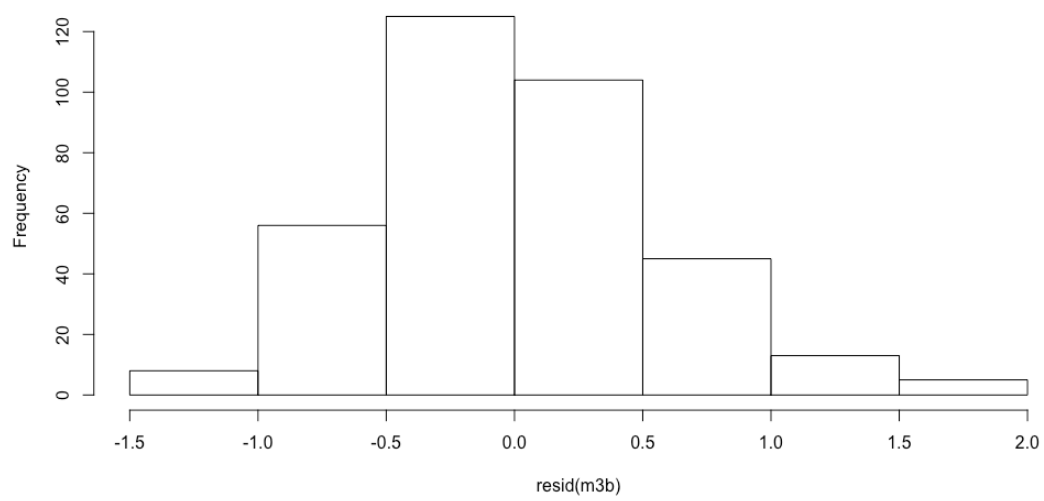




Normal Q-Q Plot



Histogram of resid(m3b)



Scaled residuals:

	Min	1Q	Median	3Q	Max
	-2.3668	-0.6731	-0.0535	0.5706	3.2077

Random effects:

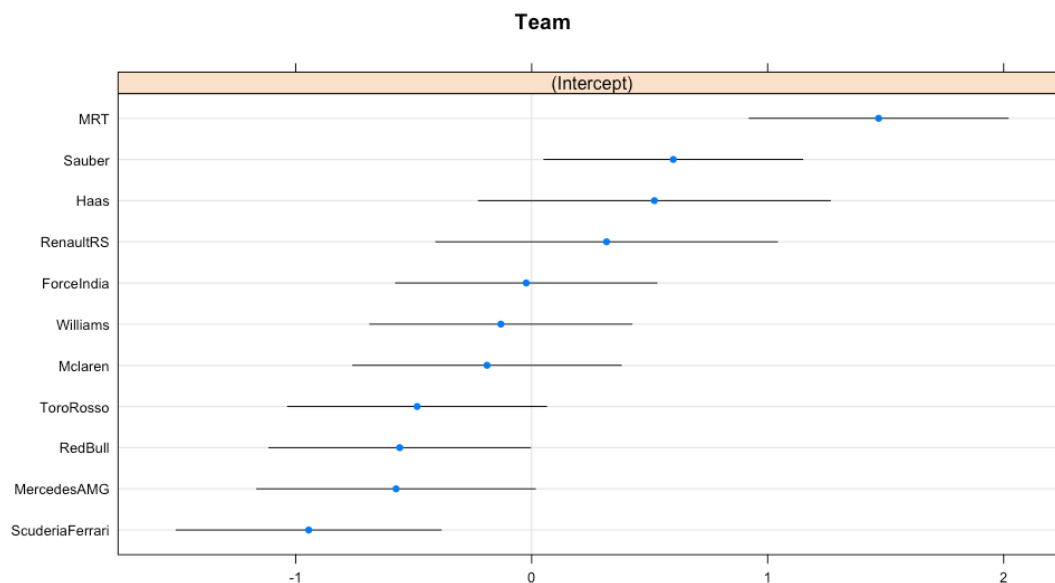
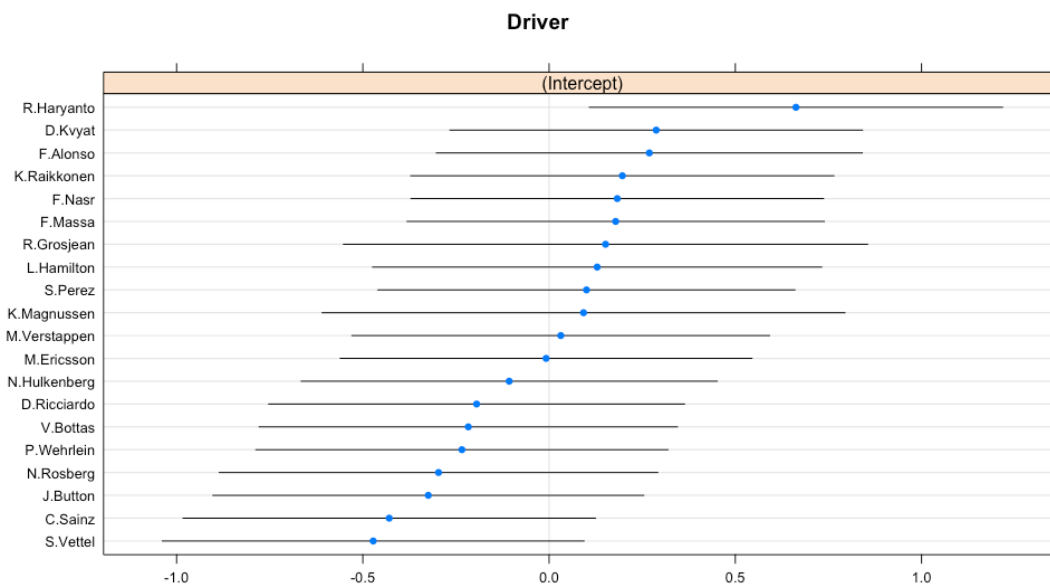
Groups	Name	Variance	Std.Dev.
Driver	(Intercept)	0.1371	0.3703
Team	(Intercept)	0.5755	0.7586
	Residual	0.3343	0.5782

Number of obs: 356, groups: Driver, 20; Team, 11

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	91.083467	0.301492	22.700000	302.109	< 2e-16 ***
nLap2	-0.007908	0.025948	333.100000	-0.305	0.76073
TyreM	-0.540036	0.203265	336.400000	-2.657	0.00826 **
TyreS	-1.259509	0.207195	334.400000	-6.079	3.30e-09 ***
nLap2:TyreM	0.071949	0.028414	335.000000	2.532	0.01179 *
nLap2:TyreS	0.191565	0.029652	332.700000	6.460	3.71e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



Annex C: Codi

```
fp <- read.csv2("FP123.csv",header=T,dec=",")

head(fp)

tail(fp)

summary(fp)

fp <- fp[,-c(8,9)]

fp$Number <- factor(fp$Number)

summary(fp)

# DESCRIPTIVA DE LES VARIABLES:

summary(fp$Driver)

sort(summary(fp$Driver))

barplot(sort(table(fp$Driver)),las=2, cex.names = 0.8, main = "Voltes per pilot",
col=c("#FFFF00","#FF9933","#0000FF","#000066","#FF3333","#000066","#660000","#FFFF00","#0000FF","#FF3333","#3399FF",
"#FF9933","#FF99CC","#FFFF00","#00FFCC","#00FFCC","#FF99CC","#3399FF","#660000","#FFFFFF","#FFFFFF"),ylim =
c(0,70))

summary(fp$Team)

sort(summary(fp$Team))

barplot(sort(table(fp$Team)),las=2, cex.names = 0.7, main = "Voltes per equips",
col=c("#FF9933","#000066","#0000FF","#FF3333","#660000","#FFFF00","#3399FF","#FF99CC","#00FFCC","#FFFFFF"),ylim =
c(0,140))

summary(fp$Engine)

sort(summary(fp$Engine))

barplot(sort(table(fp$Engine)), cex.names = 1.2, main = "Voltes per motor",
col=c("#FF9933","#FFFF00","#FF3333","#00FFCC"),ylim = c(0,350))

summary(fp$Tyre)

sort(summary(fp$Tyre))

barplot(sort(table(fp$Tyre)), cex.names = 1.5, main = "Voltes per pneumv'tic", col=c("#FF9933","#FFFF00","#FFFFFF"),ylim =
c(0,500))

summary(fp$Session)

sort(summary(fp$Session))

barplot(sort(table(fp$Session)), cex.names = 1.5, main = "Voltes per sessiv≥", col=c("#003300","#00FF00","#339966"), ylim =
c(0,500))

summary(fp$time)

h = hist(fp$time, breaks = 100, plot = F)

ccat = cut(h$breaks, c(80.21, 85.32, 91.50, 171.60))

plot(h, col=c("green","orange","red")[ccat], main = "Freqv®ncia dels temps", ylim = c(0,120), xlim = c(80,180), xlab =
"time")

# DESCRIPTIVA DE LES VARIABLES (BIVARIADA)
```



```

library('Hmisc')

## Driver

dotplot(Driver~time,data=fp,
groups=Team,col=c("#FF99CC","#660000","#FF9933","#00FFCC","#000066","#FFFF00","#3399FF","#FF3333","#0000FF","#000000"), main="Driver vs time")

## Team

dotplot(Team~time,data=fp,
groups=Team,col=c("#FF99CC","#660000","#FF9933","#00FFCC","#000066","#FFFF00","#3399FF","#FF3333","#0000FF","#000000"), main="Team vs time")

## Engine

dotplot(Engine~time,data=fp, groups=Engine, col=c("#FF3333","#FF9933","#00FFCC","#FFFF00"), main="Engine vs time")

## Tyre

dotplot(Tyre~time,data=fp, groups=Tyre, col=c("#FF9933","#000000","#FFFF00"), main="Tyre vs time")

## Session

dotplot(Session~time,data=fp, groups=Session, col=c("#00FF00","#339966","#003300"), main="Session vs time")

# DESCRIPTIVA DE LES VARIABLES (BIVARIADA) segons Tyre i Session

## Driver

dotplot(Driver~time,data=fp, groups=Session, col=c("#00FF00","#339966","#003300"), main="Driver vs time segons Session")

dotplot(Driver~time,data=fp, groups=Tyre, col=c("#FF9933","#000000","#FFFF00"), main="Driver vs time segons Tyre")

## Team

dotplot(Team~time,data=fp, groups=Session, col=c("#00FF00","#339966","#003300"), main="Team vs time segons Session")

dotplot(Team~time,data=fp, groups=Tyre, col=c("#FF9933","#000000","#FFFF00"), main="Team vs time segons Tyre")

## Engine

dotplot(Engine~time,data=fp, groups=Session, col=c("#00FF00","#339966","#003300"), main="Engine vs time segons Session")

dotplot(Engine~time,data=fp, groups=Tyre, col=c("#FF9933","#000000","#FFFF00"), main="Engine vs time segons Tyre")

## Tyre

dotplot(Tyre~time,data=fp, groups=Session, col=c("#00FF00","#339966","#003300"), main="Tyre vs time segons Session")

## Session

dotplot(Session~time,data=fp, groups=Tyre, col=c("#FF9933","#000000","#FFFF00"), main="Session vs time segons Tyre")

fp <- fp[-c(231:237),]

# S.Vandoorne

SV2 <- fp[fp$Number==2,]

FL2 <- SV2[SV2$time==min(SV2$time),]

```

```

# D.Ricciardo

DR3 <- fp[fp$Number==3,]

FL3 <- DR3[DR3$time==min(DR3$time),]

# S.Vettel

SV5 <- fp[fp$Number==5,]

FL5 <- SV5[SV5$time==min(SV5$time),]

# K.Raikkonen

KR7 <- fp[fp$Number==7,]

FL7 <- KR7[KR7$time==min(KR7$time),]

# R.Grosjean

RG8 <- fp[fp$Number==8,]

FL8 <- RG8[RG8$time==min(RG8$time),]

# M.Ericsson

ME9 <- fp[fp$Number==9,]

FL9 <- ME9[ME9$time==min(ME9$time),]

# S.Perez

SP11 <- fp[fp$Number==11,]

FL11 <- SP11[SP11$time==min(SP11$time),]

# F.Alonso

FA14 <- fp[fp$Number==14,]

FA14 <- FA14[-1,]

FL14 <- FA14[FA14$time==min(FA14$time),]

# L.Stroll

LS18 <- fp[fp$Number==18,]

FL18 <- LS18[LS18$time==min(LS18$time),]

# F.Massa

FM19 <- fp[fp$Number==19,]

FL19 <- FM19[FM19$time==min(FM19$time),]

# K.Magnussen

KM20 <- fp[fp$Number==20,]

FL20 <- KM20[KM20$time==min(KM20$time),]

# D.Kvyat

DK26 <- fp[fp$Number==26,]

```

```

FL26 <- DK26[DK26$time==min(DK26$time),]

# N.Hulkenberg
NH27 <- fp[fp$Number==27,]
FL27 <- NH27[NH27$time==min(NH27$time),]

# J.Palmer
JP30 <- fp[fp$Number==30,]
FL30 <- JP30[JP30$time==min(JP30$time),]

# E.Ocon
EO31 <- fp[fp$Number==31,]
FL31 <- EO31[EO31$time==min(EO31$time),]

#M.Verstappen
MV33 <- fp[fp$Number==33,]
FL33 <- MV33[MV33$time==min(MV33$time),]

# L.Hamilton
LH44 <- fp[fp$Number==44,]
FL44 <- LH44[LH44$time==min(LH44$time),]

# S.Sirotkin
SS46 <- fp[fp$Number==46,]
FL46 <- SS46[SS46$time==min(SS46$time),]

# C.Sainz
CS55 <- fp[fp$Number==55,]
FL55 <- CS55[CS55$time==min(CS55$time),]

# V.Bottas
VB77 <- fp[fp$Number==77,]
FL77 <- VB77[VB77$time==min(VB77$time),]

# P.Wehrlein
PW94 <- fp[fp$Number==94,]
FL94 <- PW94[PW94$time==min(PW94$time),]

# Classification
FastLaps <- rbind(FL2,FL3,FL5,FL7,FL8,FL9,FL11,FL14,FL18,FL19,FL20,FL26,FL27,FL30,FL31,FL33,FL44,FL55,FL77,FL94)

GridPosition <- (1:nrow(FastLaps))

FastLaps <- FastLaps[order(FastLaps$time),]

(FastLaps <- cbind(GridPosition,FastLaps))

```

```
# S.Vandoorne
```

```
SV2 <- fp[fp$Number==2,]
```

```
SSV2 <- SV2[SV2$Tyre=="S",]
```

```
S2 <- SSV2[SSV2$time==min(SSV2$time),]
```

```
MSV2 <- SV2[SV2$Tyre=="M",]
```

```
M2 <- MSV2[MSV2$time==min(MSV2$time),]
```

```
HSV2 <- SV2[SV2$Tyre=="H",]
```

```
H2 <- HSV2[HSV2$time==min(HSV2$time),]
```

```
SV2SMH <- rbind(S2,M2,H2)
```

```
# D.Ricciardo
```

```
DR3 <- fp[fp$Number==3,]
```

```
SDR3 <- DR3[DR3$Tyre=="S",]
```

```
S3 <- SDR3[SDR3$time==min(SDR3$time),]
```

```
MDR3 <- DR3[DR3$Tyre=="M",]
```

```
M3 <- MDR3[MDR3$time==min(MDR3$time),]
```

```
HDR3 <- DR3[DR3$Tyre=="H",]
```

```
H3 <- HDR3[HDR3$time==min(HDR3$time),]
```

```
DR3SMH <- rbind(S3,M3,H3)
```

```
# S.Vettel
```

```
SV5 <- fp[fp$Number==5,]
```

```
SSV5 <- SV5[SV5$Tyre=="S",]
```

```
S5 <- SSV5[SSV5$time==min(SSV5$time),]
```

```
MSV5 <- SV5[SV5$Tyre=="M",]
```

```
M5 <- MSV5[MSV5$time==min(MSV5$time),]
```

```
HSV5 <- SV5[SV5$Tyre=="H",]
```

```
H5 <- HSV5[HSV5$time==min(HSV5$time),]
```

```
SV5SMH <- rbind(S5,M5,H5)
```

```
# K.Raikkonen
```

```
KR7 <- fp[fp$Number==7,]
```

```
SKR7 <- KR7[KR7$Tyre=="S",]
```

```

S7 <- SKR7[SKR7$time==min(SKR7$time),]

MKR7 <- KR7[KR7$Tyre=="M",]

M7 <- MKR7[MKR7$time==min(MKR7$time),]

HKR7 <- KR7[KR7$Tyre=="H",]

H7 <- HKR7[HKR7$time==min(HKR7$time),]

KR7SMH <- rbind(S7,M7,H7)


# R.Grosjean

RG8 <- fp[fp$Number==8,]

SRG8 <- RG8[RG8$Tyre=="S",]

S8 <- SRG8[SRG8$time==min(SRG8$time),]

MRG8 <- RG8[RG8$Tyre=="M",]

M8 <- MRG8[MRG8$time==min(MRG8$time),]

HRG8 <- RG8[RG8$Tyre=="H",]

H8 <- HRG8[HRG8$time==min(HRG8$time),]

RG8SMH <- rbind(S8,M8,H8)


# M.Ericsson

ME9 <- fp[fp$Number==9,]

SME9 <- ME9[ME9$Tyre=="S",]

S9 <- SME9[SME9$time==min(SME9$time),]

MME9 <- ME9[ME9$Tyre=="M",]

M9 <- MME9[MME9$time==min(MME9$time),]

HME9 <- ME9[ME9$Tyre=="H",]

H9 <- HME9[HME9$time==min(HME9$time),]

ME9SMH <- rbind(S9,M9,H9)


# S.Perez

SP11 <- fp[fp$Number==11,]

SSP11 <- SP11[SP11$Tyre=="S",]

S11 <- SSP11[SSP11$time==min(SSP11$time),]

MSP11 <- SP11[SP11$Tyre=="M",]

M11 <- MSP11[MSP11$time==min(MSP11$time),]

```

```

HSP11 <- SP11[SP11$Tyre=="H",]
H11 <- HSP11[HSP11$time==min(HSP11$time),]
SP11SMH <- rbind(S11,M11,H11)


# F.Alonso
FA14 <- fp[fp$Number==14,]
FA14 <- FA14[-1,]
SFA14 <- FA14[FA14$Tyre=="S",]
S14 <- SFA14[SFA14$time==min(SFA14$time),]
MFA14 <- FA14[FA14$Tyre=="M",]
M14 <- MFA14[MFA14$time==min(MFA14$time),]
HFA14 <- FA14[FA14$Tyre=="H",]
H14 <- HFA14[HFA14$time==min(HFA14$time),]
FA14SMH <- rbind(S14,M14,H14)


# L.Stroll
LS18 <- fp[fp$Number==18,]
SLS18 <- LS18[LS18$Tyre=="S",]
S18 <- SLS18[SLS18$time==min(SLS18$time),]
MLS18 <- LS18[LS18$Tyre=="M",]
M18 <- MLS18[MLS18$time==min(MLS18$time),]
HLS18 <- LS18[LS18$Tyre=="H",]
H18 <- HLS18[HLS18$time==min(HLS18$time),]
LS18SMH <- rbind(S18,M18,H18)


# F.Massa
FM19 <- fp[fp$Number==19,]
SFM19 <- FM19[FM19$Tyre=="S",]
S19 <- SFM19[SFM19$time==min(SFM19$time),]
MFM19 <- FM19[FM19$Tyre=="M",]
M19 <- MFM19[MFM19$time==min(MFM19$time),]
HFM19 <- FM19[FM19$Tyre=="H",]

```

```

H19 <- HFM19[HFM19$time==min(HFM19$time),]
FM19SMH <- rbind(S19,M19,H19)

# K.Magnussen
KM20 <- fp[fp$Number==20,]
SKM20 <- KM20[KM20$Tyre=="S",]
S20 <- SKM20[SKM20$time==min(SKM20$time),]
MKM20 <- KM20[KM20$Tyre=="M",]
M20 <- MKM20[MKM20$time==min(MKM20$time),]
HKM20 <- KM20[KM20$Tyre=="H",]
H20 <- HKM20[HKM20$time==min(HKM20$time),]
KM20SMH <- rbind(S20,M20,H20)

# D.Kvyat
DK26 <- fp[fp$Number==26,]
SDK26 <- DK26[DK26$Tyre=="S",]
S26 <- SDK26[SDK26$time==min(SDK26$time),]
MDK26 <- DK26[DK26$Tyre=="M",]
M26 <- MDK26[MDK26$time==min(MDK26$time),]
HDK26 <- DK26[DK26$Tyre=="H",]
H26 <- HDK26[HDK26$time==min(HDK26$time),]
DK26SMH <- rbind(S26,M26,H26)

# N.Hulkenberg
NH27 <- fp[fp$Number==27,]
SNH27 <- NH27[NH27$Tyre=="S",]
S27 <- SNH27[SNH27$time==min(SNH27$time),]
MNH27 <- NH27[NH27$Tyre=="M",]
M27 <- MNH27[MNH27$time==min(MNH27$time),]
H27 <- MNH27[MNH27$Tyre=="H",]
H27 <- H27[H27$time==min(H27$time),]
NH27SMH <- rbind(S27,M27,H27)

```

```

# J.Palmer
JP30 <- fp[fp$Number==30,]
SJP30 <- JP30[JP30$Tyre=="S",]
S30 <- SJP30[SJP30$time==min(SJP30$time),]
MJP30 <- JP30[JP30$Tyre=="M",]
M30 <- MJP30[MJP30$time==min(MJP30$time),]
HJP30 <- JP30[JP30$Tyre=="H",]
H30 <- HJP30[HJP30$time==min(HJP30$time),]
JP30SMH <- rbind(S30,M30,H30)

# E.Ocon
EO31 <- fp[fp$Number==31,]
SEO31 <- EO31[EO31$Tyre=="S",]
S31 <- SEO31[SEO31$time==min(SEO31$time),]
MEO31 <- EO31[EO31$Tyre=="M",]
M31 <- MEO31[MEO31$time==min(MEO31$time),]
HEO31 <- EO31[EO31$Tyre=="H",]
H31 <- HEO31[HEO31$time==min(HEO31$time),]
EO31SMH <- rbind(S31,M31,H31)

# M.Verstappen
MV33 <- fp[fp$Number==33,]
SMV33 <- MV33[MV33$Tyre=="S",]
S33 <- SMV33[SMV33$time==min(SMV33$time),]
MMV33 <- MV33[MV33$Tyre=="M",]
M33 <- MMV33[MMV33$time==min(MMV33$time),]
HMV33 <- MV33[MV33$Tyre=="H",]
H33 <- HMV33[HMV33$time==min(HMV33$time),]
MV33SMH <- rbind(S33,M33,H33)

# L.Hamilton
LH44 <- fp[fp$Number==44,]
SLH44 <- LH44[LH44$Tyre=="S",]

```



```

S44 <- SLH44[SLH44$time==min(SLH44$time),]
MLH44 <- LH44[LH44$Tyre=="M",]
M44 <- MLH44[MLH44$time==min(MLH44$time),]
HLH44 <- LH44[LH44$Tyre=="H",]
H44 <- HLH44[HLH44$time==min(HLH44$time),]
LH44SMH <- rbind(S44,M44,H44)

# C.Sainz
CS55 <- fp[fp$Number==55,]
SCS55 <- CS55[CS55$Tyre=="S",]
S55 <- SCS55[SCS55$time==min(SCS55$time),]
MCS55 <- CS55[CS55$Tyre=="M",]
M55 <- MCS55[MCS55$time==min(MCS55$time),]
HCS55 <- CS55[CS55$Tyre=="H",]
H55 <- HCS55[HCS55$time==min(HCS55$time),]
CS55SMH <- rbind(S55,M55,H55)

# V.Bottas
VB77 <- fp[fp$Number==77,]
SVB77 <- VB77[VB77$Tyre=="S",]
S77 <- SVB77[SVB77$time==min(SVB77$time),]
MVB77 <- VB77[VB77$Tyre=="M",]
M77 <- MVB77[MVB77$time==min(MVB77$time),]
HVB77 <- VB77[VB77$Tyre=="H",]
H77 <- HVB77[HVB77$time==min(HVB77$time),]
VB77SMH <- rbind(S77,M77,H77)

# P.Wehrlein
PW94 <- fp[fp$Number==94,]
SPW94 <- PW94[PW94$Tyre=="S",]
S94 <- SPW94[SPW94$time==min(SPW94$time),]
MPW94 <- PW94[PW94$Tyre=="M",]
M94 <- MPW94[MPW94$time==min(MPW94$time),]

```

```

HPW94 <- PW94[PW94$Tyre=="H",]

H94 <- HPW94[HPW94$time==min(HPW94$time),]

PW94SMH <- rbind(S94,M94,H94)

FastLapsSMH <-
rbind(SV2SMH,DR3SMH,SV5SMH,KR7SMH,RG8SMH,ME9SMH,SP11SMH,FA14SMH,LS18SMH,FM19SMH,KM20SMH,DK26S
MH,NH27SMH,JP30SMH,EO31SMH,MV33SMH,LH44SMH,CS55SMH,VB77SMH,PW94SMH)

dotplot(time~Tyre|Driver,FastLapsSMH, col=c("#FFFF00","#000000","#FF9933"))

xyplot(time~Tyre|Driver,FastLapsSMH,type=c("r","p"),col=c("#FFFF00","#000000","#FF9933"))


# Classification segons Tyre

# Soft (S)

FastLapsS <- rbind(FL2,FL3,FL5,FL7,FL8,FL9,FL11,FL14,FL18,FL19,FL20,FL26,FL27,FL30,FL31,FL33,FL44,FL55,FL77,FL94)

GridPosition <- (1:nrow(FastLapsS))

FastLapsS <- FastLapsS[order(FastLapsS$time),]

FastLapsS <- cbind(GridPosition,FastLapsS)

BesttimeSoft <- FastLapsS$Driver


# Medium (M)

FastLapsM <- rbind(M2,M3,M5,M7,M8,M9,M11,M14,M18,M19,M20,M26,M27,M30,M31,M33,M44,M55,M77,M94)

FastLapsM <- FastLapsM[order(FastLapsM$time),]

FastLapsM <- cbind(GridPosition,FastLapsM)

BesttimeMedium <- FastLapsM$Driver


# Hard (H)

FastLapsH <- rbind(H2,H3,H5,H7,H8,H9,H11,H14,H18,H19,H20,H26,H27,H30,H31,H33,H44,H55,H77,H94)

FastLapsH <- FastLapsH[order(FastLapsH$time),]

GridPositionH <- (1:nrow(FastLapsH))

FastLapsH <- cbind(GridPositionH,FastLapsH)

BesttimeHard <- FastLapsH$Driver

DSM <- data.frame(GridPosition,BesttimeSoft,BesttimeMedium)

add.col<-function(df, BesttimeHard) {n.row<-dim(df)[1]

length(BesttimeHard)<-n.row

cbind(df, BesttimeHard)

}

```

```

(BesttimeSMH<-add.col(DSM,BesttimeHard))

library(car)

# S.Vandoorne

SVP2 <- SV2[SV2$Session=="FP2",]

Lap <- (1:nrow(SVP2))

SVP2 <- cbind(Lap,SVP2)

(scatterplot(time ~ Lap | Tyre, data=SVP2, xlab="Lap", ylab="time", main="S.Vandoorne FP2 times",
col=c('white','#000000','#FFCC33'), labels=row.names(SVP2)))

SVP2 <- SVP2[-c(1:7,11,12,15,18,21),]

scatterplot(time ~ Lap | Tyre, data=SVP2, xlab="Lap", ylab="time", main="S.Vandoorne Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(SVP2))


# D.Ricciardo

DRP3 <- DR3[DR3$Session=="FP2",]

Lap <- (1:nrow(DRP3))

DRP3 <- cbind(Lap,DRP3)

(scatterplot(time ~ Lap | Tyre, data=DRP3, xlab="Lap", ylab="time", main="D.Ricciardo FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(DRP3)))

DRP3 <- DRP3[-c(1,6:8,11,12,17),]

scatterplot(time ~ Lap | Tyre, data=DRP3, xlab="Lap", ylab="time", main="D.Ricciardo Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(DRP3))


# S.Vettel

SVP5 <- SV5[SV5$Session=="FP2",]

Lap <- (1:nrow(SVP5))

SVP5 <- cbind(Lap,SVP5)

(scatterplot(time ~ Lap | Tyre, data=SVP5, xlab="Lap", ylab="time", main="S.Vettel FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(SVP5)))

SVP5 <- SVP5[-c(1:7,14,17:20,22),]

scatterplot(time ~ Lap | Tyre, data=SVP5, xlab="Lap", ylab="time", main="S.Vettel Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(SVP5))


# K.Raikkonen

KRP7 <- KR7[KR7$Session=="FP2",]

Lap <- (1:nrow(KRP7))

```

```

KRP7 <- cbind(Lap,KRP7)

(scatterplot(time ~ Lap | Tyre, data=KRP7, xlab="Lap", ylab="time", main="K.Raikkonen FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(KRP7)))

KRP7 <- KRP7[-c(1:8,10,11,13),]

scatterplot(time ~ Lap | Tyre, data=KRP7, xlab="Lap", ylab="time", main="K.Raikkonen Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(KRP7))


# R.Grosjean

RGP8 <- RG8[RG8$Session=="FP2",]

Lap <- (1:nrow(RGP8))

RGP8 <- cbind(Lap,RGP8)

(scatterplot(time ~ Lap | Tyre, data=RGP8, xlab="Lap", ylab="time", main="R.Grosjean FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(RGP8)))

RGP8 <- RGP8[-c(8:12,23),]

scatterplot(time ~ Lap | Tyre, data=RGP8, xlab="Lap", ylab="time", main="R.Grosjean Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(RGP8))


# M.Ericsson

MEP9 <- ME9[ME9$Session=="FP2",]

Lap <- (1:nrow(MEP9))

MEP9 <- cbind(Lap,MEP9)

(scatterplot(time ~ Lap | Tyre, data=MEP9, xlab="Lap", ylab="time", main="M.Ericsson FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(MEP9)))

MEP9 <- MEP9[-c(2,4,6,8,10:13,16,21),]

scatterplot(time ~ Lap | Tyre, data=MEP9, xlab="Lap", ylab="time", main="M.Ericsson Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(MEP9))


# S.Perez

SPP11 <- SP11[SP11$Session=="FP2",]

Lap <- (1:nrow(SPP11))

SPP11 <- cbind(Lap,SPP11)

(scatterplot(time ~ Lap | Tyre, data=SPP11, xlab="Lap", ylab="time", main="S.Perez FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(SPP11)))

SPP11 <- SPP11[-c(2,5,7:9),]

scatterplot(time ~ Lap | Tyre, data=SPP11, xlab="Lap", ylab="time", main="S.Perez Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(SPP11))

```

```

# F.Alonso

FAP14 <- FA14[FA14$Session=="FP2",]

Lap <- (1:nrow(FAP14))

FAP14 <- cbind(Lap,FAP14)

(scatterplot(time ~ Lap | Tyre, data=FAP14, xlab="Lap", ylab="time", main="F.Alonso FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(FAP14)))

FAP14 <- FAP14[-c(2,4:6,9),]

scatterplot(time ~ Lap | Tyre, data=FAP14, xlab="Lap", ylab="time", main="F.Alonso Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(FAP14))


# L.Stroll

LSP18 <- LS18[LS18$Session=="FP2",]

Lap <- (1:nrow(LSP18))

LSP18 <- cbind(Lap,LSP18)

(scatterplot(time ~ Lap | Tyre, data=LSP18, xlab="Lap", ylab="time", main="L.Stroll FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(LSP18)))

LSP18 <- LSP18[-c(3,6,8:10,18:20),]

scatterplot(time ~ Lap | Tyre, data=LSP18, xlab="Lap", ylab="time", main="L.Stroll Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(LSP18))


# F.Massa

FMP19 <- FM19[FM19$Session=="FP2",]

Lap <- (1:nrow(FMP19))

FMP19 <- cbind(Lap,FMP19)

(scatterplot(time ~ Lap | Tyre, data=FMP19, xlab="Lap", ylab="time", main="F.Massa FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(FMP19)))

FMP19 <- FMP19[-c(2,4,6:8,11,13,24),]

scatterplot(time ~ Lap | Tyre, data=FMP19, xlab="Lap", ylab="time", main="F.Massa Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(FMP19))


# K.Magnussen

KMP20 <- KM20[KM20$Session=="FP2",]

Lap <- (1:nrow(KMP20))

KMP20 <- cbind(Lap,KMP20)

(scatterplot(time ~ Lap | Tyre, data=KMP20, xlab="Lap", ylab="time", main="K.Magnussen FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(KMP20)))

KMP20 <- KMP20[-c(1,4,7,9,11,12),]

```

```
scatterplot(time ~ Lap | Tyre, data=KMP20, xlab="Lap", ylab="time", main="K.Magnussen Real Race Pace",col=c('white','#000000','#FFCC33'), labels=row.names(KMP20))
```

```
# D.Kvyat
```

```
DKP26 <- DK26[DK26$Session=="FP2",]
```

```
Lap <- (1:nrow(DKP26))
```

```
DKP26 <- cbind(Lap,DKP26)
```

```
(scatterplot(time ~ Lap | Tyre, data=DKP26, xlab="Lap", ylab="time", main="D.Kvyat FP2 times",col=c('white','#000000','#FFCC33'), labels=row.names(DKP26)))
```

```
DKP26 <- DKP26[-c(2,4,5,12),]
```

```
scatterplot(time ~ Lap | Tyre, data=DKP26, xlab="Lap", ylab="time", main="D.Kvyat Real Race Pace",col=c('white','#000000','#FFCC33'), labels=row.names(DKP26))
```

```
# N.Hulkenberg
```

```
NHP27 <- NH27[NH27$Session=="FP2",]
```

```
Lap <- (1:nrow(NHP27))
```

```
NHP27 <- cbind(Lap,NHP27)
```

```
(scatterplot(time ~ Lap | Tyre, data=NHP27, xlab="Lap", ylab="time", main="N.Hulkenberg FP2 times",col=c('white','#000000','#FFCC33'), labels=row.names(NHP27)))
```

```
NHP27 <- NHP27[-c(2,4:8,14,19,23),]
```

```
scatterplot(time ~ Lap | Tyre, data=NHP27, xlab="Lap", ylab="time", main="N.Hulkenberg Real Race Pace",col=c('white','#000000','#FFCC33'), labels=row.names(NHP27))
```

```
# J.Palmer
```

```
JPP30 <- JP30[JP30$Session=="FP2",]
```

```
Lap <- (1:nrow(JPP30))
```

```
JPP30 <- cbind(Lap,JPP30)
```

```
(scatterplot(time ~ Lap | Tyre, data=JPP30, xlab="Lap", ylab="time", main="J.Palmer FP2 times",col=c('white','#000000','#FFCC33'), labels=row.names(JPP30)))
```

```
JPP30 <- JPP30[-c(2,4,6,8,10:12,14,29),]
```

```
scatterplot(time ~ Lap | Tyre, data=JPP30, xlab="Lap", ylab="time", main="J.Palmer Real Race Pace",col=c('white','#000000','#FFCC33'), labels=row.names(JPP30))
```

```
# E.Ocon
```

```
EOP31 <- EO31[EO31$Session=="FP2",]
```

```
Lap <- (1:nrow(EOP31))
```

```
EOP31 <- cbind(Lap,EOP31)
```

```

(scatterplot(time ~ Lap | Tyre, data=EOP31, xlab="Lap", ylab="time", main="E.Ocon FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(EOP31)))

EOP31 <- EOP31[-c(2,5:9),]

scatterplot(time ~ Lap | Tyre, data=EOP31, xlab="Lap", ylab="time", main="E.Ocon Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(EOP31))


# M.Verstappen

MVP33 <- MV33[MV33$Session=="FP2",]

Lap <- (1:nrow(MVP33))

MVP33 <- cbind(Lap,MVP33)

(scatterplot(time ~ Lap | Tyre, data=MVP33, xlab="Lap", ylab="time", main="M.Verstappen FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(MVP33)))

MVP33 <- MVP33[-c(2:4,6,7,9,11,13),]

scatterplot(time ~ Lap | Tyre, data=MVP33, xlab="Lap", ylab="time", main="M.Verstappen Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(MVP33))


# L.Hamilton

LHP44 <- LH44[LH44$Session=="FP2",]

Lap <- (1:nrow(LHP44))

LHP44 <- cbind(Lap,LHP44)

(scatterplot(time ~ Lap | Tyre, data=LHP44, xlab="Lap", ylab="time", main="L.Hamilton FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(LHP44)))

LHP44 <- LHP44[-c(1,2,4,6:9,15,18,21:23,25),]

scatterplot(time ~ Lap | Tyre, data=LHP44, xlab="Lap", ylab="time", main="L.Hamilton Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(LHP44))


# C.Sainz

CSP55 <- CS55[CS55$Session=="FP2",]

Lap <- (1:nrow(CSP55))

CSP55 <- cbind(Lap,CSP55)

(scatterplot(time ~ Lap | Tyre, data=CSP55, xlab="Lap", ylab="time", main="C.Sainz FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(CSP55)))

CSP55 <- CSP55[-c(1,3,5,6,13,22),]

scatterplot(time ~ Lap | Tyre, data=CSP55, xlab="Lap", ylab="time", main="C.Sainz Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(CSP55))

```

```

# V.Bottas

VBP77 <- VBP77[VB77$Session=="FP2",]

Lap <- (1:nrow(VBP77))

VBP77 <- cbind(Lap,VBP77)

(scatterplot(time ~ Lap | Tyre, data=VBP77, xlab="Lap", ylab="time", main="V.Bottas FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(VBP77)))

VBP77 <- VBP77[-c(2,4,7,15,16,21,22,26),]

scatterplot(time ~ Lap | Tyre, data=VBP77, xlab="Lap", ylab="time", main="V.Bottas Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(VBP77))


# P.Wehrlein

PWP94 <- PWP94[PWP94$Session=="FP2",]

Lap <- (1:nrow(PWP94))

PWP94 <- cbind(Lap,PWP94)

(scatterplot(time ~ Lap | Tyre, data=PWP94, xlab="Lap", ylab="time", main="P.Wehrlein FP2
times",col=c('white','#000000','#FFCC33'), labels=row.names(PWP94)))

PWP94 <- PWP94[-c(2,3,5,7,10,11),]

scatterplot(time ~ Lap | Tyre, data=PWP94, xlab="Lap", ylab="time", main="P.Wehrlein Real Race
Pace",col=c('white','#000000','#FFCC33'), labels=row.names(PWP94))

summary(aov(time~Team,FastLaps))

summary(modT<-lm(time~Team,FastLaps))

summary(aov(time~Engine,FastLaps))

summary(lm(time~Engine,FastLaps))

FMHR <- FastLaps[c(1:4,7,10,11,19),]

summary(aov(time~Engine,FMHR))

summary(modE<-lm(time~Engine,FMHR))

summary(aov(time~Tyre,FastLapsSMH))

summary(modTy<-lm(time~Tyre,FastLapsSMH))

summary(aov(time~Session,fp))

summary(modS<-lm(time~Session,fp))

mean(fp$time[fp$Session=="FP2"]) < mean(fp$time[fp$Session=="FP3"])

qfC <- read.csv2("Qualifying.csv",header=T,dec=",")

head(qfC)

data.frame(GridPosition,BesttimeSoft,qfC$Driver) # Tots els pilots van fer la volta mes rapida amb Soft.

DiferenciaCat <- qfC$FinalLap-qfC$FPBestLap

```



```

DifCat <- data.frame(qfC$Driver,qfC$Team,qfC$FinalLap,qfC$FPBestLap, DiferenciaCat)

library(formattable)

plain_formatter <- formatter("span")

plain_formatter(c(1, 1, 1))

width_formatter <- formatter("span",

style = x ~ style(width = suffix(x, "px")))

width_formatter(c(1,1,1))

sign_formatter <- formatter("span", style = x ~ style(color = ifelse(x > 0, "red", "green")),

x ~ icontext(ifelse(x < 0, "arrow-down", "arrow-up"), formattable(x, digits = 3, format = "f")))

sign_formatter(c(-1, 0, 1))

formattable(DifCat, list(DiferenciaCat = sign_formatter))

formattable(DifCat[with(DifCat,order(DiferenciaCat)),], list(DiferenciaCat = sign_formatter)) # Show in the New Window but
sort by improve

Dif <- round(DiferenciaCat / (qfC$FPBestLap/100),3)

ImproCat <- data.frame(qfC$Driver,qfC$Team,Dif)

formattable(ImproCat[with(ImproCat,order(Dif)),], list(Dif = sign_formatter))

summary(aov(FPBestLap~FinalLap,qfC))

data.frame(GridPosition,qfC$Driver,qfC$FinalLap,qfC$KPH)

attach(qfC)

plot(qfC$FinalLap,qfC$KPH)

scatterplot(FinalLap ~ KPH | Engine, data=qfC,

          xlab="Speed", ylab="time",

          main="Speed vs. time",

          labels=row.names(qfC), col=c("#FF3333","#FF9933","#00FFCC","#FFFF00"),smoother=F)


pstCat <- read.csv2("PitStop.csv",header=T,dec=",")

dotplot(Driver~Duration,data=pstCat,main="Pit Stops segons Driver")

pstCat[pstCat[, "Duration"] > 40,]

pstCat[pstCat[, "Lap"] == 1,]

pstCat <- pstCat[-c(2),]

mpstCat <- aggregate(pstCat$Duration, by=list(Category=pstCat$Team), FUN=mean)

colnames(mpstCat) <- c("Team","Duration")

(mpstCat <- mpstCat[order(mpstCat$Duration), ])

```

```

dotplot(Team~Duration,data=mpstCat,main="Mitjana dels Pit Stops segons
Team",col=c("#00FFCC","#000066","#0000FF","#000000","#FF3333","#FFF000","#FF9933","#3399FF","#660000","#FF99CC"
))

install.packages("plotly")

library(plotly)

library(ggplot2)

library(gapminder)

df <- read.csv2("Race.csv",header=T,dec=",")

p <- df %>%

  plot_ly(

    x = ~nLap,

    y = ~Ttime,

    color = ~Team,

    frame = ~nLap,

    text = ~Driver,

    hoverinfo = "text",

    type = 'scatter',

    mode = 'markers'

  ) %>%

  layout()

p

library(lme4)

library(lmerTest)

fp2 <- read.csv2("FP123b.csv",header=T,dec=",")

head(fp2)

library(lattice)

xyplot(time~nLap | Driver:Session,fp2,ylim=c(60,120),type=c("g","p","r"),group=Tyre,auto.key = T)

summary(m1<-lmer(time~Session*nLap+Engine+Tyre+(1 | Team/Driver),fp2))

plot(m1)

qqnorm(resid(m1))

hist(resid(m1))

boxplot(time~Tyre,fp2)

library(car)

carr=split(fp2, list(fp2$Driver,fp2$Session))

```

```

temps=lapply(carr,function(el)

cbind(el$Tyre[-1]!=el$Tyre[-nrow(el)],el$time[-1]/el$time[-nrow(el)]-1,el[-1,]))

separa = function(aux,minvol=5,cvmax=0.1){

val = NULL

if (nrow(aux) > 1){

canvis = c(1,which(aux[1:(nrow(aux)-1),1]==1),nrow(aux)+1)

sec = cbind(canvis[-length(canvis)],canvis[-1]-1)


k = 1

for(i in 1:nrow(sec)){

dad = aux[sec[i,1]:sec[i,2],,drop=F]

if (sec[i,2]-sec[i,1]+1>=minvol){

out = Boxplot(dad[, "time"])

if (length(out)>1) resul<-dad[-out,] else resul<-dad

cv = sd(resul[, "time"])/mean(resul[, "time"])

if (cv<cvmax & nrow(resul)>=minvol) {

val[[k]] = resul

k = k+1

}

}

}

}

val

}

library(plyr)

tandes=lapply(temps,separa,minvol=5,cv=0.04)

tandes=compact(tandes)

tandes=lapply(tandes,function(el) lapply(el,function(el1) cbind(el1,nLap2=1:nrow(el1))))

tandes2=do.call(rbind,unlist(tandes,recursive=F)) # Unifiquem la llista de llistes en una per tots

xyplot(time~nLap|Driver:Session,tandes2,ylim=c(60,120),type=c("g","p","r"),group=Tyre,auto.key = T)

xyplot(time~nLap2|Driver:Session,tandes2,ylim=c(60,120),type=c("g","p","r"),group=Tyre,auto.key = T)

```

```

summary(m1<-lmer(time~Session*nLap+Engine+Tyre+(1|Team/Driver),tandes2))

summary(m1<-lmer(time~Session*nLap2+Engine+Tyre+(1|Team/Driver),tandes2))

plot(m1)

qqnorm(resid(m1))

hist(resid(m1))

tandesFP2=tandes2[tandes2$Session=="FP2",]

xyplot(time~nLap2|Driver,tandesFP2,ylim=c(60,120),type=c("g","p","r"),group=Tyre,auto.key = T)

summary(m2<-lmer(time~nLap2+Engine+Tyre+(1|Team)+(1|Driver),tandesFP2))

plot(m2,id=0.003)

aa=names(sort(abs(resid(m2)),decreasing=T))[1:13]

dadesFin=tandesFP2[!row.names(tandesFP2)%in%aa,]

m3a<-lmer(time~nLap2*Tyre+Engine+(1|Team)+(1|Driver),dadesFin)

summary(m3a)

m3b<-lmer(time~nLap2*Tyre+(1|Team)+(1|Driver),dadesFin)

summary(m3b)

m3<-lmer(time~nLap2+Tyre+(1|Team)+(1|Driver),dadesFin)

summary(m3)

AIC(m3a)

AIC(m3b)

AIC(m3)

xyplot(time~nLap2|Driver,dadesFin,ylim=c(60,120),type=c("g","p","r"),group=Tyre,auto.key = T)

plot(m3,id=0.003)

qqnorm(resid(m3))

hist(resid(m3))

ranef(m3)

n = 1:66

time <- (88.263+0.122*n)*n+(87.236+0.122*(66-n))*(66-n)

minim <- min(time)

match(minim, time)

# 31 amb M

# 35 amb S

temps = minim + 22.5

h=temps[1]%/%3600

```

```

m=(temps[1]-h*3600)%/%60
s=temps[1]-h*3600-m*60
paste0(h,".",m,".",s)
xyplot(time~nLap2 | Driver,dadesFin,ylim=c(60,120),type=c("g","p","r"),group=Tyre,auto.key = T)
library(lattice)
dotplot(ranef(m3,condVar=T))
detach(package:lmerTest)
detach(package:lme4)
library(nlme)
dades=tandesFP2[!row.names(tandesFP2)%in%aa,c(4,5,6,7,12,14)]
dades=groupedData(time~nLap2 | Team/Driver,dades)
summary(m1b<-lme(time~nLap2+Tyre,dades,random=~1 | Team/Driver))
ranef(m1b)
plot(augPred(m1b,level=0:2),ylim=c(75,95))
(pr1<-coef(m1b)[order(coef(m1b)[,1]),1:2])
coef(m3)
library(data.table)
noudades=cbind(unique(dadesFin[,4:5]),nLap2=rep(1,20),Tyre=rep("S",20))
sort(predict(m3,newdata=noudades))
volta=63
daux=unique(dadesFin[,4:5])
noudades=data.frame(Driver=rep(daux[,1],volta),Team=rep(daux[,2],volta),nLap2=rep(1:volta,rep(20,volta)),Tyre=rep("S",20*volta))
voltes=matrix(predict(m3,newdata=noudades),nr=20)
row.names(voltes)=daux[,1]
pr2=apply(t(voltes),2,sum)
(temps=sort(pr2))
h=temps[1]%/%3600
m=(temps[1]-h*3600)%/%60
s=temps[1]-h*3600-m*60
paste0(h,".",m,".",s)
qu=read.csv2("qualifying.csv")
qu$FinalLap
row.names(qu)=paste0(qu$Team,"/",qu$Driver)

```

```

qu=cbind(qu,pred=pr1[row.names(qu),1])
plot(FPBestLap~FinalLap,qu,type="n")
text(qu$FinalLap,qu$FPBestLap,qu$Driver,cex=0.7)
qu$FPBestLap=qu$FPBestLap-min(qu$FPBestLap)
qu$Obs=qu$FinalLap-min(qu$FinalLap)
plot(obs~FPBestLap,qu,pch="+")
text(qu$FPBestLap,qu$Obs+0.1,qu$Driver,cex=0.7)
abline(c(0,1))

(rmse=sqrt(mean((qu$FPBestLap-qu$Obs)^2)))
(mae=mean(abs(qu$FPBestLap-qu$Obs)))
qu$FinalLap
row.names(qu)=paste0(qu$Team,"/",qu$Driver)
qu=cbind(qu,pred=pr1[row.names(qu),1])
plot(pred~FinalLap,qu,type="n")
text(qu$FinalLap,qu$pred,qu$Driver,cex=0.7)
qu$predic=qu$pred-min(qu$pred)
qu$Obs=qu$FinalLap-min(qu$FinalLap)
plot(obs~predic,qu,pch="+")
text(qu$predic,qu$Obs+0.1,qu$Driver,cex=0.7)
abline(c(0,1))

(rmse=sqrt(mean((qu$predic-qu$Obs)^2)))
(mae=mean(abs(qu$predic-qu$Obs)))
volta=63
daux=unique(dadesFin[,4:5])
noudades=data.frame(Driver=rep(daux[,1],volta),Team=rep(daux[,2],volta),nLap2=rep(1:volta,rep(20,volta)),Tyre=rep("S",20*volta))
voltes=matrix(predict(m3,newdata=noudades),nr=20)
row.names(voltes)=daux[,1]
pr2=apply(t(voltes),2,sum)
(temps=sort(pr2))
ra=read.csv2("race.csv")
ra=ra[ra$nLap==volta,]
ra=cbind(ra,pred=pr2[as.character(ra$Driver)])
plot(Ttime~pred,ra,pch="+")

```

```
text(ra$pred,ra$Ttime+volta/10,ra$Driver,cex=0.7)
abline(c(0,1))
(data.frame(ra$Driver,ra$pred))
(data.frame(ra$Driver,ra$Ttime))
ra$predic=ra$pred-min(ra$pred)
ra$obs=ra$Ttime-min(ra$Ttime)
plot(obs~predic,ra,pch="+",main=paste0("Volta ",volta))
text(ra$predic,ra$obs+volta/10,ra$Driver,cex=0.7)
(rmse=sqrt(mean((ra$predic-ra$obs)^2)))
(mae=mean(abs(ra$predic-ra$obs)))
```